GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 15:20:40 ; Search time 3001 Seconds
(without alignments)
13731.947 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1307	92.3	1385	9	AF069987	AF069987 Homo sapi
2		571	40.3	4079	9	AF069984	AF069984 Homo sapi
3	c	571	40.3	167863	2	AL590651	AL590651 Homo sapi
4		571	40.3	200822	9	AL591806	AL591806 Human DNA
5	c	326	23.0	377	6	AX368386	AX368386 Sequence
6	c	173	12.2	441	11	G13436	G13436 human STS W
7	c	87	6.1	2261	6	AX397469	AX397469 Sequence
8	c	87	6.1	2261	9	AK001497	AK001497 Homo sapi
9		47	3.3	847	11	G72919	G72919 MARC 2849-2
10		33	2.3	1338	10	AF069988	AF069988 Mus muscu
11		33	2.3	1365	10	BC021634	BC021634 Mus muscu
12		33	2.3	4481	10	AF069985	AF069985 Mus muscu
13		33	2.3	181583	2	AC105589	AC105589 Rattus no
14		33	2.3	211772	10	AC084821	AC084821 Mus muscu
15	c	33	2.3	215043	10	AC087229	AC087229 Mus muscu
16	c	32	2.3	115667	2	AC125857	AC125857 Rattus no
17		31	2.2	144986	2	AC108571	AC108571 Rattus no
18		30	2.1	1587	9	BC017413	BC017413 Homo sapi
19		30	2.1	2421	10	BC016425	BC016425 Mus muscu
20		30	2.1	3300	6	AR212407	AR212407 Sequence
21		30	2.1	4113	10	AF230376	AF230376 Meriones
22	c	30	2.1	44645	9	AC108058	AC108058 Homo sapi
23		30	2.1	60773	2	AC069576	AC069576 Homo sapi
24	c	30	2.1	252059	2	AC102081	AC102081 Mus muscu
25		29	2.0	345	10	RNRIPRL38	X57007 Rat mRNA fo
26		29	2.0	549	9	AF112344	AF112344 Homo sapi
27	c	29	2.0	564	6	AX407609	AX407609 Sequence
28		29	2.0	876	5	AF175983	AF175983 Rana sylv
29		29	2.0	1049	8	AF071889	AF071889 Prunus ar
30		29	2.0	1075	9	BC014385	BC014385 Homo sapi
31		29	2.0	1134	10	AF058799	AF058799 Mus muscu
32		29	2.0	1193	6	A29423	A29423 putative bo
33		29	2.0	1193	6	AR022483	AR022483 Sequence
34		29	2.0	1193	6	AR068961	AR068961 Sequence
35		29	2.0	1193	6	AR103330	AR103330 Sequence
36		29	2.0	1193	6	AR134988	AR134988 Sequence
37		29	2.0	1193	6	AR141817	AR141817 Sequence
38		29	2.0	1193	6	AR143334	AR143334 Sequence
39		29	2.0	1193	6	AR151943	AR151943 Sequence
40		29	2.0	1193	6	I36423	I36423 Sequence 13
41		29	2.0	1193	6	I88123	I88123 Sequence 13
42		29	2.0	1211	5	AY065841	AY065841 Danio rer
43		29	2.0	1267	10	BC004064	BC004064 Mus muscu
44		29	2.0	1488	9	HSU79258	U79258 Human clone
45		29	2.0	1667	3	AY047568	AY047568 Drosophil

ALIGNMENTS

RESULT 1
AF069987
LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.
ACCESSION AF069987
VERSION AF069987.1 GI:3228665
KEYWORDS .
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1385)
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

```
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            1. .1385
                     /gene="NIT1"
     CDS             77. .1060
                     /gene="NIT1"
                     /note="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC39907.1"
                     /db_xref="GI:3228666"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
BASE COUNT      322 a    382 c    357 g    324 t
ORIGIN

  Query Match             92.3%;  Score 1307;  DB 9;  Length 1385;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1307;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy     170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy     230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy     290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318

Qy     350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy     410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy     470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy     530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy     590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 618

Qy     650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     619 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 678

Qy     710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     679 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 738

Qy     770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     739 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 798

Qy     830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     799 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 858

Qy     890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     859 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 918

Qy     950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     919 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 978

Qy    1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     979 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1038

Qy    1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1039 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1098

Qy    1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1099 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1158

Qy    1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1159 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1218

Qy    1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1219 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1278

Qy    1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1279 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1338

Qy    1370 GAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1339 GAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1385


RESULT 2
AF069984
LOCUS       AF069984                4079 bp    DNA     linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069984
VERSION     AF069984.1  GI:3242977
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
```

```
          Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE 98337986
   PUBMED 9671749
REFERENCE 2  (bases 1 to 4079)
  AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
          Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
          Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE   Direct Submission
  JOURNAL Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
          Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            635. .3569
                     /gene="NIT1"
     mRNA            join(635. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3569)
                     /gene="NIT1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            635. .680
                     /gene="NIT1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(679. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3267)
                     /gene="NIT1"
                     /note="Nit1; alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC39901.1"
                     /db_xref="GI:3242978"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
     exon            681. .711
                     /gene="NIT1"
                     /note="1A; alternatively spliced"
     exon            712. .741
                     /gene="NIT1"
                     /note="1B; alternatively spliced"
     exon            940. .1053
                     /gene="NIT1"
                     /note="1C; alternatively spliced"
     exon            1288. .1383
                     /gene="NIT1"
                     /number=2
     exon            1636. .1890
                     /gene="NIT1"
                     /number=3
     exon            2011. .2114
                     /gene="NIT1"
                     /number=4
     exon            2321. .2454
                     /gene="NIT1"
                     /number=5
     exon            2609. .2734
                     /gene="NIT1"
                     /number=6
     exon            3001. .3569
                     /gene="NIT1"
                     /number=7
BASE COUNT      940 a   1082 c   1043 g   1012 t      2 others
ORIGIN

  Query Match           40.3%;  Score 571;  DB 9;  Length 4079;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2999 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 3058

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3059 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 3118

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3119 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 3178

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3179 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 3238

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3239 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 3298

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3299 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 3358

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3359 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 3418

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3419 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 3478

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3479 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 3538

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||||
Db    3539 TGGCATTGAAAAATATAATAATCATAAAGTC 3569


RESULT 3
AL590651/c
LOCUS       AL590651              167863 bp    DNA     linear   HTG 18-AUG-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL590651
VERSION     AL590651.4  GI:13990236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167863)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13751001.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA137A12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
```

```
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 164729 bases at least Q40
          Consensus quality: 165486 bases at least Q30
          Consensus quality: 166169 bases at least Q20
          Insert size: 167063; sum-of-contigs
          Insert size: 171878; 3.4% error; agarose-fp
          Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
          coverage: 8.11x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     9373: contig of 9373 bp in length
          *     9374 9473: gap of      100 bp
          *     9474    29891: contig of 20418 bp in length
          *    29892 29991: gap of      100 bp
          *    29992    36239: contig of 6248 bp in length
          *    36240 36339: gap of      100 bp
          *    36340   103784: contig of 67445 bp in length
          *   103785 103884: gap of      100 bp
          *   103885   114878: contig of 10994 bp in length
          *   114879 114978: gap of      100 bp
          *   114979   126347: contig of 11369 bp in length
          *   126348 126447: gap of      100 bp
          *   126448   137764: contig of 11317 bp in length
          *   137765 137864: gap of      100 bp
          *   137865   163160: contig of 25296 bp in length
          *   163161 163260: gap of      100 bp
          *   163261   167863: contig of 4603 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-137A12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .9373
                     /note="assembly_fragment:02360
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9474. .29891
                     /note="assembly_fragment:01854
                     fragment_chain:1"
     misc_feature    29992. .36239
                     /note="assembly_fragment:00805
                     fragment_chain:1"
     misc_feature    36340. .103784
                     /note="assembly_fragment:01467
                     fragment_chain:1"
     misc_feature    103885. .114878
                     /note="assembly_fragment:00991
                     fragment_chain:2"
     misc_feature    114979. .126347
                     /note="assembly_fragment:01503
                     fragment_chain:2"
     misc_feature    126448. .137764
                     /note="assembly_fragment:01297
                     fragment_chain:2"
     misc_feature    137865. .163160
                     /note="assembly_fragment:00117
                     fragment_chain:2"
     misc_feature    163261. .167863
                     /note="assembly_fragment:01267
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    43927 a  39935 c  40002 g  43197 t    802 others
```

```
ORIGIN

  Query Match             40.3%;  Score 571;  DB 2;  Length 167863;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149682 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 149623

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149622 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 149563

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149562 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 149503

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149502 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 149443

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149442 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 149383

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149382 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 149323

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149322 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 149263

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149262 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 149203

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149202 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 149143

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||
Db 149142 TGGCATTGAAAAATATAATAATCATAAAGTC 149112


RESULT 4
AL591806
LOCUS       AL591806              200822 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
            sequence.
ACCESSION   AL591806
VERSION     AL591806.16  GI:18476709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200822)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17902927.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
```

```
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-544M22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-544M22 The true
            left end of clone RP11-137A12 is at 156538 in this sequence. The
            true right end of clone RP11-381D2 is at 145015 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .200822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-544M22"
                     /clone_lib="RPCI-11.2"
     misc_feature    10395. .10397
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    44098. .44107
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    59585. .59604
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    90797. .90883
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    90861. .90883
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    55215 a  49052 c  46183 g  50372 t
ORIGIN

  Query Match             40.3%;  Score 571;  DB 9;  Length 200822;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174499 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 174558

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174559 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 174618

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174619 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 174678

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174679 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 174738

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174739 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 174798

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174799 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 174858

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174859 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 174918

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174919 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 174978

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174979 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 175038

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||
Db 175039 TGGCATTGAAAAATATAATAATCATAAAGTC 175069


RESULT 5
AX368386/c
LOCUS       AX368386                377 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1096 from Patent WO0204514.
ACCESSION   AX368386
VERSION     AX368386.1  GI:18856459
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1096 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       94 a    100 c    112 g     71 t
ORIGIN

  Query Match             23.0%;  Score 326;  DB 6;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-190;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy     189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy     249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy     309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy     369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy     429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18
```

```
Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1


RESULT 6
G13436/c
LOCUS       G13436                   441 bp    DNA     linear   STS 04-JUN-1996
DEFINITION  human STS WI-12338, sequence tagged site.
ACCESSION   G13436
VERSION     G13436.1  GI:1127545
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TATATTTTCAATGCCAGTGCTG
            Primer B: AAACTTTCACCTGAGCTTCACC
            STS size: 130
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T78621).
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="750.8 cR from top of Chr1 linkage group"
     STS             16. .145
     primer_bind     16. .38
     primer_bind     complement(124. .145)
BASE COUNT      107 a     95 c    130 g    103 t      6 others
ORIGIN

  Query Match             12.2%;  Score 173;  DB 11;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 2e-95;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1178 CCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCT 1237
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    216 CCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCT 157

Qy   1238 GCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 1297
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 GCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 97

Qy   1298 TGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGG 1350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 TGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGG 44


RESULT 7
AX397469/c
LOCUS       AX397469                 2261 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1684 from Patent WO0212328.
ACCESSION   AX397469
VERSION     AX397469.1  GI:21068216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   King,G.E., Meagher,M.J., Xu,J. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0212328-A 1684 14-FEB-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2261
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      543 a    594 c    541 g    583 t
ORIGIN

  Query Match             6.1%;  Score 87;  DB 6;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-42;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 8
AK001497/c
LOCUS       AK001497                 2261 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ10635 fis, clone NT2RP2005669, highly similar
            to Homo sapiens mRNA for DEDD protein.
ACCESSION   AK001497
VERSION     AK001497.1  GI:7022790
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2005669.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
            Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,
            Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K.,
            Masuho,Y. and Kanehori,K.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2261)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
```

```
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2261
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2005669"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3~mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
BASE COUNT      543 a    594 c    541 g    583 t
ORIGIN

  Query Match             6.1%;  Score 87;  DB 9;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-42;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 9
G72919
LOCUS       G72919                   847 bp    DNA     linear   STS 08-AUG-2001
DEFINITION  MARC 2849-2850:991933517:1 SCF - porcine spleen Sus scrofa STS
            genomic, sequence tagged site.
ACCESSION   G72919
VERSION     G72919.1  GI:15146949
KEYWORDS    STS.
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 847)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
            Keele,J.W.
  TITLE     Single nucleotide polymorphism (SNP) discovery in expressed porcine
            genes
  JOURNAL   Unpublished
COMMENT
            Contact: Freking BA
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4278
            Fax: 402 762 4173
            Email: freking@email.marc.usda.gov
            Primer A: GGGTTCAGCTTTTGGATCTG
            Primer B: GCTGAGCCTGTGTTTCATCA
            STS size: 700
            PCR Profile:
                  Hotstart:  95 degrees for 15 minutes
                  Denature:  95 degrees for 30 seconds
                  Anneal:    58 degrees
                  Extension: 68 degrees for 2 minutes
                  Cycles:    32 to 45
            Protocol:
                  Template:       50-200 ng genomic DNA
                  Primer:         each 20 pmoles
                  dNTPs:          each 88 uM
                  Taq Polymerase: 0.25 units (Qiagen HotStar)

            Buffer:
                  Commercially supplied Qiagen HotStar buffer

            The STS is derived from PCR amplicons generated from genomic DNA,
            sequenced from each end using the amplification primers.  The
            sequence does not necessarily represent the entire amplicon.
            Sequence derived from PolyPhred was trimmed from each end of each
            unique contig until five consecutive bases exceeded a quality score
            threshold of 20, and the next 10 bases averaged a quality score of
            20 or greater. Amplicon size was estimated by agarose gel
            electrophoresis.
FEATURES             Location/Qualifiers
     source          1. .847
                     /organism="Sus scrofa"
                     /strain="white composite, duroc, meishan, minzhu,
                     fengjing, crossbreds"
                     /db_xref="taxon:9823"
                     /sex="male and female"
                     /clone_lib="SCF - porcine spleen"
                     /dev_stage="adult"
                     /note="Organ: spleen"
     STS             <1. .>847
BASE COUNT      205 a    249 c    200 g    193 t
ORIGIN

  Query Match             3.3%;  Score 47;  DB 11;  Length 847;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    873 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCA 919
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCA 578


RESULT 10
AF069988
LOCUS       AF069988                 1338 bp    mRNA    linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase 1 (Nit1) mRNA, complete cds.
ACCESSION   AF069988
VERSION     AF069988.1  GI:3228667
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1338
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            1. .1338
                     /gene="Nit1"
     CDS             58. .1029
```

```
                     /gene="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC40185.1"
                     /db_xref="GI:3228668"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      347 a    335 c    330 g    326 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 1338;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db    1269 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1301


RESULT 11
BC021634
LOCUS       BC021634               1365 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
            complete cds.
ACCESSION   BC021634
VERSION     BC021634.1  GI:18204912
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6754855.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:13825 IMAGE:4008543"
                     /tissue_type="Mammary tumor metastatized to lung.
                     MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
                     enhancer."
                     /clone_lib="NCI_CGAP_Lu30"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             59. .1030
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAH21634.1"
                     /db_xref="GI:18204913"
                     /db_xref="LocusID:27045"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYSSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      368 a    334 c    331 g    332 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 1365;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db    1270 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1302


RESULT 12
AF069985
LOCUS       AF069985               4481 bp    DNA     linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase homolog 1 (Nit1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069985
VERSION     AF069985.1  GI:3242979
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4481
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            541. .3944
                     /gene="Nit1"
     mRNA            join(541. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3944)
                     /gene="Nit1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            541. .606
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(605. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3658)
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
```

```
                     /protein_id="AAC40184.1"
                     /db_xref="GI:3242980"
                     /translation="MLGFITRPPHQLLCTGYRLLRIPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
     exon            863. .986
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=2
     exon            1177. .1263
                     /gene="Nit1"
                     /number=3
     exon            1638. .1889
                     /gene="Nit1"
                     /number=4
     exon            2015. .2118
                     /gene="Nit1"
                     /number=5
     exon            2362. .2495
                     /gene="Nit1"
                     /number=6
     exon            2626. .2751
                     /gene="Nit1"
                     /number=7
     exon            3392. .3944
                     /gene="Nit1"
                     /number=8
BASE COUNT     1101 a   1132 c   1133 g   1115 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 4481;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db   3898 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 3930


RESULT 13
AC105589
LOCUS       AC105589               181583 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
            ***, 67 unordered pieces.
ACCESSION   AC105589
VERSION     AC105589.2  GI:21736456
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 181583)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18092812.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNGI
                Center clone name: CH230-242B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 112684 bases at least Q40
                Consensus quality: 120012 bases at least Q30
                Consensus quality: 127587 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 67 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1312: contig of 1312 bp in length
            *     1313     1412: gap of unknown length
            *     1413     2762: contig of 1350 bp in length
            *     2763     2862: gap of unknown length
            *     2863     4412: contig of 1550 bp in length
            *     4413     4512: gap of unknown length
            *     4513     5532: contig of 1020 bp in length
            *     5533     5632: gap of unknown length
            *     5633     6814: contig of 1182 bp in length
```

```
*     6815     6914: gap of unknown length
*     6915     8099: contig of 1185 bp in length
*     8100     8199: gap of unknown length
*     8200     9786: contig of 1587 bp in length
*     9787     9886: gap of unknown length
*     9887    11574: contig of 1688 bp in length
*    11575    11674: gap of unknown length
*    11675    12816: contig of 1142 bp in length
*    12817    12916: gap of unknown length
*    12917    14067: contig of 1151 bp in length
*    14068    14167: gap of unknown length
*    14168    15370: contig of 1203 bp in length
*    15371    15470: gap of unknown length
*    15471    16926: contig of 1456 bp in length
*    16927    17026: gap of unknown length
*    17027    18088: contig of 1062 bp in length
*    18089    18188: gap of unknown length
*    18189    19315: contig of 1127 bp in length
*    19316    19415: gap of unknown length
*    19416    20772: contig of 1357 bp in length
*    20773    20872: gap of unknown length
*    20873    21937: contig of 1065 bp in length
*    21938    22037: gap of unknown length
*    22038    23147: contig of 1110 bp in length
*    23148    23247: gap of unknown length
*    23248    24803: contig of 1556 bp in length
*    24804    24903: gap of unknown length
*    24904    26111: contig of 1208 bp in length
*    26112    26211: gap of unknown length
*    26212    27647: contig of 1436 bp in length
*    27648    27747: gap of unknown length
*    27748    30257: contig of 2510 bp in length
*    30258    30357: gap of unknown length
*    30358    31725: contig of 1368 bp in length
*    31726    31825: gap of unknown length
*    31826    32912: contig of 1087 bp in length
*    32913    33012: gap of unknown length
*    33013    34593: contig of 1581 bp in length
*    34594    34693: gap of unknown length
*    34694    36338: contig of 1645 bp in length
*    36339    36438: gap of unknown length
*    36439    38139: contig of 1701 bp in length
*    38140    38239: gap of unknown length
*    38240    39349: contig of 1110 bp in length
*    39350    39449: gap of unknown length
*    39450    40968: contig of 1519 bp in length
*    40969    41068: gap of unknown length
*    41069    42068: contig of 1000 bp in length
*    42069    42168: gap of unknown length
*    42169    43869: contig of 1701 bp in length
*    43870    43969: gap of unknown length
*    43970    46141: contig of 2172 bp in length
*    46142    46241: gap of unknown length
*    46242    48158: contig of 1917 bp in length
*    48159    48258: gap of unknown length
*    48259    50613: contig of 2355 bp in length
*    50614    50713: gap of unknown length
*    50714    52725: contig of 2012 bp in length
*    52726    52825: gap of unknown length
*    52826    54534: contig of 1709 bp in length
*    54535    54634: gap of unknown length
*    54635    56455: contig of 1821 bp in length
*    56456    56555: gap of unknown length
*    56556    58494: contig of 1939 bp in length
*    58495    58594: gap of unknown length
*    58595    60496: contig of 1902 bp in length
*    60497    60596: gap of unknown length
*    60597    62628: contig of 2032 bp in length
*    62629    62728: gap of unknown length
*    62729    65060: contig of 2332 bp in length
*    65061    65160: gap of unknown length
*    65161    66543: contig of 1383 bp in length
*    66544    66643: gap of unknown length
*    66644    68412: contig of 1769 bp in length
*    68413    68512: gap of unknown length
*    68513    71585: contig of 3073 bp in length
*    71586    71685: gap of unknown length
*    71686    74575: contig of 2890 bp in length
*    74576    74675: gap of unknown length
*    74676    77865: contig of 3190 bp in length
*    77866    77965: gap of unknown length
*    77966    80442: contig of 2477 bp in length
*    80443    80542: gap of unknown length
*    80543    83108: contig of 2566 bp in length
*    83109    83208: gap of unknown length
*    83209    86895: contig of 3687 bp in length
*    86896    86995: gap of unknown length
*    86996    91038: contig of 4043 bp in length
*    91039    91138: gap of unknown length
*    91139    94412: contig of 3274 bp in length
*    94413    94512: gap of unknown length
*    94513    96959: contig of 2447 bp in length
*    96960    97059: gap of unknown length
*    97060    98988: contig of 1929 bp in length
*    98989    99088: gap of unknown length
*    99089   101633: contig of 2545 bp in length
*   101634   101733: gap of unknown length
*   101734   105899: contig of 4166 bp in length

  Query Match             2.3%;  Score 33;  DB 2;  Length 181583;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db  135691 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 135723


RESULT 14
AC084821
LOCUS       AC084821               211772 bp    DNA     linear   ROD 19-FEB-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-395h6, complete sequence.
ACCESSION   AC084821
VERSION     AC084821.25  GI:18702388
KEYWORDS    HTG.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-395h6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 19, 2002 this sequence version replaced gi:18390260.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------
FEATURES             Location/Qualifiers
     source          1. .211772
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
```

```
                     /clone="rp23-395h6"
                     /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT    55607 a  48743 c  49203 g  58215 t      4 others
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 211772;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db    3237 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 3269


RESULT 15
AC087229/c
LOCUS       AC087229              215043 bp    DNA     linear   ROD 14-AUG-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,
            complete sequence.
ACCESSION   AC087229
VERSION     AC087229.24  GI:22094398
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Mus musculus Chromosome 1 BAC Clone rp23-191a19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Aug 3, 2002 this sequence version replaced gi:22038565.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------
FEATURES             Location/Qualifiers
     source          1. .215043
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-191a19"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT    55806 a  49340 c  51162 g  58732 t      3 others
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 215043;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db   87663 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 87631


Search completed: December 14, 2002, 18:13:49
Job time : 4176 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 13:59:05 ; Search time 263 Seconds
(without alignments)
12124.827 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1314	92.8	1416	21	AAZ46101	Nucleotide sequenc
	2	1160	81.9	1382	21	AAF16257	Human prostate can
	3	766	54.1	1203	20	AAX30398	DNA encoding a hum
c	4	326	23.0	377	24	ABK39058	cDNA encoding lung
	5	91	6.4	505	22	AAH07371	Human cDNA clone (
	6	87	6.1	592	22	AAH11324	Human cDNA clone (
c	7	87	6.1	2261	22	AAH14367	Human cDNA sequenc
c	8	87	6.1	2261	24	ABK46133	cDNA encoding colo
	9	58	4.1	60	24	ABN39628	Human spliced tran
	10	52	3.7	539	22	AAK61236	Human immune/haema
	11	30	2.1	3300	20	AAZ11915	Human potassium ch
	12	29	2.0	80	24	ABA82722	Human protective D
	13	29	2.0	223	23	ABV08728	Human prostate exp
c	14	29	2.0	311	24	ABL82990	Human ovarian canc
c	15	29	2.0	384	23	ABV13242	Human prostate exp
c	16	29	2.0	406	23	ABV04073	Human prostate exp
c	17	29	2.0	416	23	ABV34361	Human prostate exp
	18	29	2.0	417	23	ABV58938	Human prostate exp
	19	29	2.0	528	23	ABV38626	Human prostate exp
	20	29	2.0	563	20	AAX98006	Human secreted pro
c	21	29	2.0	564	24	ABN93758	Gene #256 used to
c	22	29	2.0	748	22	AAK63693	Human immune/haema
	23	29	2.0	807	24	ABA82714	Human protective D
	24	29	2.0	1193	13	AAQ30670	GGF2BPP2.CDS. Syn
	25	29	2.0	1193	15	AAQ58303	GGF-II cDNA sequen
	26	29	2.0	1193	15	AAQ62849	GGF-II cDNA sequen
	27	29	2.0	1193	16	AAQ74885	Putative bovine gl
	28	29	2.0	1193	17	AAT48079	Bovine neuregulin
	29	29	2.0	1193	17	AAT30997	Bovine glial growt
	30	29	2.0	1193	17	AAT06703	Bovine glial growt
	31	29	2.0	1211	22	AAH31352	Human secreted pro
	32	29	2.0	1608	21	AAC95501	Human secreted pro
	33	29	2.0	2317	22	AAS03917	Human secreted pro
	34	29	2.0	3625	23	AAF87120	NOV9 coding sequen
	35	29	2.0	4843	20	AAX37553	Human HLIG-1 cDNA.
c	36	28	2.0	319	24	ABL79372	Human ovarian canc
	37	28	2.0	324	23	ABV49621	Human prostate exp
	38	28	2.0	326	23	ABV19858	Human prostate exp
	39	28	2.0	428	23	ABV17820	Human prostate exp
	40	28	2.0	469	23	ABV47612	Human prostate exp
c	41	28	2.0	619	22	AAL22931	Human breast cance
c	42	28	2.0	711	22	AAI97858	Human neuroblastom
	43	28	2.0	941	24	ABN98591	Arabidopsis thalia
	44	28	2.0	1108	13	AAQ30671	GGF2BPP3.CDS. Syn
	45	28	2.0	1108	15	AAQ58304	GGF-II cDNA sequen

ALIGNMENTS

```
RESULT 1
AAZ46101
ID   AAZ46101 standard; DNA; 1416 BP.
XX
AC   AAZ46101;
XX
DT   05-MAY-2000  (first entry)
XX
DE   Nucleotide sequence of the coding region of NIT1 gene.
XX
KW   NIT1 gene; nitrilase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW   FRA3B; cancer; genome allele inactivation; ss.
XX
OS   Homo sapiens.
OS   Mus sp.
OS   Drosophila melanogaster.
OS   Caenorhabditis elegans.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1415
FT                   /*tag= a
FT                   /product= (pos: 18..20, aa: Xaa)
FT                   /product= (pos: 1179..1181, aa: Gln)
FT                   /product= (pos: 1182..1184, aa: Ala)
FT   .               /note= "contains 7 internal stop codons; Xaa is an
FT                         unspecified amino acid"
XX
PN   WO200003685-A2.
XX
PD   27-JAN-2000.
XX
PF   20-JUL-1999;   99WO-US16366.
```

```
XX
PR   20-JUL-1998;   98US-0093350.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce CM;
XX
DR   WPI; 2000-171195/15.
DR   P-PSDB; AAY68739.
XX
PT   Novel nitrilase homologs used as diagnostic and therapeutic reagents
PT   for the detection and treatment of cancer  -
XX
PS   Claim 6; Fig 6; 25pp; English.
XX
CC   The present sequence represents the coding region of human, murine,
CC   Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC   human and mouse NIT1 genes are members of an uncharacterised
CC   mammalian gene family with homology to bacterial and plant nitrilases.
CC   The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC   for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC   In mouse and humans, FHIT and NIT are encoded by two different genes,
CC   localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC   human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC   chromosomal fragile site FRA3B, is often altered in most common forms
CC   of human cancer. The Nit1 protein overcomes the mutated inactivation
CC   of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC   and analogues of them, and antibodies are used as diagnostic and
CC   therapeutic reagents for the detection and treatment of cancers.
XX
SQ   Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

  Query Match             92.8%;  Score 1314;  DB 21;  Length 1416;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1414;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCACTCGCTGCGGCCTNTCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCCCACTCGCTGCGGCCTNTCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGG 60

Qy      61 TTTTGGCTATATCTTCATGTAGGACCTACTCCCTATCCCGTCGGCCGCGGCTGGGCTTCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTTGGCTATATCTTCATGTAGGACCTACTCCCTATCCCGTCGGCCGCGGCTGGGCTTCA 120

Qy     121 TCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAAC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAAC 180

Qy     181 TCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCG 240

Qy     241 AACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTA 300

Qy     301 AAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360

Qy     361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCAC 420

Qy     421 TGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGT 480

Qy     481 CCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATT 540

Qy     541 GTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGT 600

Qy     601 GTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660

Qy     661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATG 720

Qy     721 ACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCT 780

Qy     781 ATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCC 840

Qy     841 GTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGA 900

Qy     901 AGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

Qy     961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGC 1020

Qy    1021 GCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACC 1080

Qy    1081 CACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCC 1140

Qy    1141 ACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTG 1200
           ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db    1141 ACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCACCCACAGCTCCCCTCACTTG 1200

Qy    1201 GAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCT 1260

Qy    1261 GAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTT 1320

Qy    1321 TATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAA 1380

Qy    1381 TAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||||||||
Db    1381 TAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416


RESULT 2
AAF16257
ID   AAF16257 standard; cDNA; 1382 BP.
XX
AC   AAF16257;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
```

```
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB57054.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 1124; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

  Query Match             81.9%;  Score 1160;  DB 21;  Length 1382;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1280;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy     170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy     230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy     290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318

Qy     350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy     410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy     470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy     530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy     590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGG-CTATGTGTGAAAGCAACTCTAC 617

Qy     650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     618 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 677

Qy     710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 737

Qy     770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 797

Qy     830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 857

Qy     890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 917

Qy     950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 977

Qy    1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1037

Qy    1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1097

Qy    1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1157

Qy    1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1217

Qy    1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1277

Qy    1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1337

Qy    1370 GAAAAATATAATAATCATAAA 1390
           ||||||||||||||||||||
Db    1338 GAAAAATATAATAATCATAAA 1358


RESULT 3
AAX30398
ID   AAX30398 standard; DNA; 1203 BP.
XX
AC   AAX30398;
XX
DT   14-MAY-1999  (first entry)
XX
```

```
DE   DNA encoding a human secreted protein.
XX
KW   Secreted protein; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder;
KW   CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW   renal disease; diabetes; inflammation; allergy; ischemic shock;
KW   Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW   prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9907891-A1.
XX
PD   18-FEB-1999.
XX
PF   04-AUG-1998;   98WO-US16235.
XX
PR   19-AUG-1997;   97US-0056732.
PR   05-AUG-1997;   97US-0054798.
PR   05-AUG-1997;   97US-0054803.
PR   05-AUG-1997;   97US-0054804.
PR   05-AUG-1997;   97US-0054806.
PR   05-AUG-1997;   97US-0054807.
PR   05-AUG-1997;   97US-0054808.
PR   05-AUG-1997;   97US-0054809.
PR   05-AUG-1997;   97US-0055309.
PR   05-AUG-1997;   97US-0055310.
PR   05-AUG-1997;   97US-0055312.
PR   05-AUG-1997;   97US-0055386.
PR   05-AUG-1997;   97US-0055311.
PR   18-AUG-1997;   97US-0055970.
PR   18-AUG-1997;   97US-0055986.
PR   19-AUG-1997;   97US-0056365.
PR   19-AUG-1997;   97US-0056366.
PR   19-AUG-1997;   97US-0056557.
PR   19-AUG-1997;   97US-0056370.
PR   19-AUG-1997;   97US-0056371.
PR   19-AUG-1997;   97US-0056563.
PR   19-AUG-1997;   97US-0056731.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Ebner R,  Ferrie AM,  Greene JM,  Janat F,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR,  Young PE,  Yu G;
XX
DR   WPI; 1999-167452/14.
DR   P-PSDB; AAY10877.
XX
PT   New isolated human genes encoding secreted polypeptides - useful for
PT   diagnosis and treatment of pathalogical diseases
XX
PS   Claim 3; Page 265-266; 331pp; English.
XX
CC   The specification describes secreted proteins and their corresponding
CC   polynucleotides which are useful for preventing, treating or ameliorating
CC   medical conditions, e.g. by protein or gene therapy. Pathological
CC   conditions can also be diagnosed by determining the amount of the
CC   secreted polypeptides in a sample or by determining the presence of
CC   mutations in the polynucleotides. Specific uses are described for each
CC   of the products, based on which tissues they are most highly
CC   expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC   and cognitive disorders, schizophrenia, cardiovascular disorders,
CC   prostate diseases, asthma, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC   thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
XX
SQ   Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;
```

```
  Query Match             54.1%;  Score 766;  DB 20;  Length 1203;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-305;
  Matches  766;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     628 CTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACAC 687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 CTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACAC 220

Qy     688 CAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGG 747
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 CAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGG 280

Qy     748 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 807
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 340

Qy     808 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAG 867
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAG 400

Qy     868 TGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGG 927
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 TGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGG 460

Qy     928 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTG 987
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTG 520

Qy     988 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACC 1047
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACC 580

Qy    1048 GCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGA 1107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 GCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGA 640

Qy    1108 GTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTG 1167
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 GTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTG 700

Qy    1168 GAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 1227
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     701 GAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 760

Qy    1228 CAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGT 1287
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     761 CAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGT 820

Qy    1288 TTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTG 1347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     821 TTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTG 880

Qy    1348 AGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||||||||||||||||||
Db     881 AGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 926
```

```
RESULT 4
ABK39058/c
ID   ABK39058 standard; cDNA; 377 BP.
XX
AC   ABK39058;
XX
DT   21-MAY-2002  (first entry)
XX
DE   cDNA encoding lung tumour protein clone R0130:B11.
XX
KW   Lung tumour; cancer; T cell; immune response stimulator;
KW   cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200204514-A2.
```

```
XX
PD   17-JAN-2002.
XX
PF   10-JUL-2001; 2001WO-US22058.
XX
PR   11-JUL-2000; 2000US-0614124.
PR   29-AUG-2000; 2000US-0651563.
PR   08-SEP-2000; 2000US-0658824.
PR   26-SEP-2000; 2000US-0671325.
PR   06-OCT-2000; 2000US-0677419.
PR   30-OCT-2000; 2000US-0702705.
PR   13-DEC-2000; 2000US-0736457.
PR   03-MAY-2001; 2001US-0849626.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
PI   Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
PI   Mcnabb A, Wang A, Fanger N, Switzer A, Mcneill PD, Clapper JD;
XX
DR   WPI; 2002-164634/21.
XX
PT   Novel polynucleotide encoding a lung tumour polypeptide useful for
PT   stimulating and/or expanding T cells specific for a tumour protein -
XX
PS   Claim 1; SEQ ID No 1096; 223pp; English.
XX
CC   The invention describes an isolated polynucleotide and polypeptide
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein for determining the presence of a cancer in a patient. A
CC   composition containing the polynucleotide and/or polypeptide is useful
CC   for treating a lung cancer in a patient. The polypeptide is useful for
CC   removing tumour cells from a biological sample. The polynucleotide is
CC   also useful as probe or primer to detect the level of mRNA encoding a
CC   tumour protein. This sequence encodes a lung tumour associated protein
CC   or protein fragment, described in the method of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 377 BP; 94 A; 100 C; 112 G; 71 T; 0 other;

  Query Match             23.0%;  Score 326;  DB 24;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 3.7e-124;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy     189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy     249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy     309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy     369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy     429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy     489 GGTTTCCATGAGCGTGG 505
           |||||||||||||||||
Db      17 GGTTTCCATGAGCGTGG 1


RESULT 5
AAH07371
ID   AAH07371 standard; cDNA; 505 BP.
XX
AC   AAH07371;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:4206.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 4206; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 505 BP; 105 A; 120 C; 167 G; 108 T; 5 other;

  Query Match              6.4%;  Score 91;  DB 22;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 1e-27;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCATG 79
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCATG 78

Qy      80 TAGGACCTACTCCCTATCCCGTCGGCCGCGG 110
           ||||||||||||||||||||||||||||||
Db      79 TAGGACCTACTCCCTATCCCGTCGGCCGCGG 109


RESULT 6
AAH11324
ID   AAH11324 standard; cDNA; 592 BP.
XX
AC   AAH11324;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:8159.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 8159; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 592 BP; 164 A; 134 C; 138 G; 148 T; 8 other;

  Query Match             6.1%;  Score 87;  DB 22;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 60

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db     61 ATTGAAAAATATAATAATCATAAAGTC 87


RESULT 7
AAH14367/c
ID   AAH14367 standard; cDNA; 2261 BP.
XX
AC   AAH14367;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11772.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11772; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
```

```
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

  Query Match             6.1%;  Score 87;  DB 22;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 8
ABK46133/c
ID   ABK46133 standard; cDNA; 2261 BP.
XX
AC   ABK46133;
XX
DT   05-JUN-2002  (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 1684.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic;
KW   immunotherapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US24218.
XX
PR   03-AUG-2000; 2000US-223283P.
PR   28-MAR-2001; 2001US-279763P.
PR   29-JUN-2001; 2001US-302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE,  Meagher MJ,  Xu J,  Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers
PT   for the progression of cancer -
XX
PS   Claim 1; SEQ ID No 1684; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer.
CC   ABK44450-ABK46237 represent coding sequences of human colon tumour
CC   proteins of the invention.
CC   Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC   for this patent did not form part of the printed specification but was
CC   supplied by the European Patent Office.
XX
SQ   Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

  Query Match             6.1%;  Score 87;  DB 24;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 9
ABN39628
ID   ABN39628 standard; DNA; 60 BP.
XX
AC   ABN39628;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human spliced transcript detection oligonucleotide SEQ ID NO:12376.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB01903.
XX
PR   28-JUL-2000; 2000US-221607P.
PR   02-MAY-2001; 2001US-287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A,  Wasserman A,  Mintz E,  Mintz L,  Faigler S;
XX
DR   WPI; 2002-257383/30.
XX
PT   New oligonucleotide libraries comprising oligonucleotides which
PT   selectively hybridize to mRNAs transcribed from a transcription unit of
PT   a genome, useful for detecting tissue-, pathology-, and
PT   developmental-specific genes -
XX
PS   Example 1; SEQ ID 12376; 47pp; English.
XX
CC   The present invention describes oligonucleotide libraries for detecting
CC   messenger RNAs that populate a (sub-)transcriptome, where the
CC   (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC   transcription units that populate a genome. The library comprises
CC   several oligonucleotides, each capable of hybridising selectively to a
CC   set of messenger RNAs transcribed from a given transcription unit of
CC   the genome, which encodes one or more messenger RNA splice variants.
CC   The oligonucleotide libraries are useful for detecting mRNAs from a
CC   biological sample, in expression profiling studies, in qualitatively or
CC   quantitatively characterising the corresponding transcriptome, and in
CC   detecting RNA transcripts and splice variants of human or animal
CC   transcriptomes. The libraries may also be used as specialised mini
CC   libraries to detect transcripts of a sub-transcriptome under a
CC   particular biological or pathological state, and so allowing the
CC   detection of tissue- and pathology-specific genes such as those genes
CC   only expressed in specific tissue under a specific pathological
CC   condition; to detect developmental specific genes; and to detect RNA
CC   transcripts and splice variants of a transcriptome of a patient suffering
CC   from a particular disorder. ABN27253 to ABN59589 represent
CC   oligonucleotide sequences from rats, humans and mice, which are used in
CC   the exemplification of the present invention.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence 60 BP; 9 A; 20 C; 14 G; 17 T; 0 other;

  Query Match             4.1%;  Score 58;  DB 24;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   58;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCA 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCA 60


RESULT 10
AAK61236
ID   AAK61236 standard; cDNA; 539 BP.
XX
AC   AAK61236;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6296.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
```

```
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM88455.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 1; SEQ ID NO 6296; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

  Query Match             3.7%;  Score 52;  DB 22;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 1e-11;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     109 GGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 GGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 422


RESULT 11
AAZ11915
ID   AAZ11915 standard; cDNA; 3300 BP.
XX
AC   AAZ11915;
XX
DT   30-NOV-1999  (first entry)
XX
DE   Human potassium channel K+Hnov59 cDNA.
XX
KW   Potassium channel; ataxia; arrhythmia; epilepsy; Bartter's syndrome;
KW   cardiovascular disorder; CNS disorder; renal disorder; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             50..1285
FT                   /*tag=  a
FT                   /product= "Human K+Hnov59 potassium channel"
XX
PN   WO9943696-A1.
XX
PD   02-SEP-1999.
XX
PF   22-FEB-1999;   99WO-US03826.
XX
PR   19-JAN-1999;   99US-0116448.
PR   25-FEB-1998;   98US-0076687.
PR   07-AUG-1998;   98US-0095836.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Curran ME,  Hu P,  Miller AP,  Rutter M,  Wang J;
XX
DR   WPI; 1999-527591/44.
DR   P-PSDB; AAY34133.
XX
PT   New nucleic acids encoding mammalian K+Hnov potassium channel
PT   proteins, useful for the diagnosis and treatment of episodic ataxia
PT   with myokymia, cardiac arrhythmia, epilepsy and Bartter's syndrome
XX
PS   Claim 4; Page 102-104; 112pp; English.
XX
CC   This sequence represents human potassium channel K+Hnov59 cDNA.
CC   K+Hnov proteins have a high degree of homology to known potassium
CC   channels and may be alpha subunits, which form the functional channel,
CC   or accessory subunits that act to modulate the channel activity. K+Hnov59
CC   is a 4 transmembrane domain, 2 pore domain potassium channel. The gene
CC   is located on chromosome 19, determined via PCR chromosomal
CC   localisation using primers AAZ11939 and AAZ11940. K+Hnov cDNAs
CC   were isolated by extension of expressed sequence tags (ESTs) which were
CC   related but not identical to known human potassium channels. Potential
CC   polymorphisms detected as sequence variants between multiple
CC   independent clones. Potassium channels have critical roles in various
CC   cell types and biochemical pathways. Defective potassium channels are
CC   known to cause four human diseases: episodic ataxia with myokymia;
CC   cardiac arrhythmia (long QT syndrome); epilepsy; and Bartter's syndrome.
CC   As potassium channels are critical components of virtually all cells,
CC   it is likely that abnormal potassium channels are also implicated in
CC   certain renal, cardiovascular and central nervous system (CNS)
CC   disorders. Nucleotides encoding K+Hnov proteins may be used for
CC   identifying homologous or related proteins and the DNA sequences encoding
CC   them. They may be used to produce compositions that modulate the
CC   expression and function of the K+Hnov protein and in studying the
CC   biochemical pathways associated with it. They may also be used for the
CC   recombinant production of K+Hnov protein in fermentation cultures.
CC   Additionally, such nucleotides may be used in gene therapy protocols for
CC   the treatment of diseases associated with abnormal potassium channels.
XX
SQ   Sequence 3300 BP; 997 A; 629 C; 680 G; 994 T; 0 other;

  Query Match             2.1%;  Score 30;  DB 20;  Length 3300;
  Best Local Similarity   100.0%;  Pred. No. 0.0093;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||
Db    3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 3267
```

```
RESULT 12
ABA82722
ID   ABA82722 standard; DNA; 80 BP.
XX
AC   ABA82722;
XX
DT   07-FEB-2002  (first entry)
XX
DE   Human protective DNA sequence CNI-00735 fragment #8.
XX
KW   Human; protective sequence; cell death; cancer; autoimmune disease;
KW   neurological disorder; stroke; cytostatic; neuroprotective; gene therapy;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200176457-A2.
XX
PD   18-OCT-2001.
XX
PF   09-APR-2001; 2001WO-US11663.
XX
PR   11-APR-2000; 2000US-0547735.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Thomas MB,  Portbury SD,  Puranam K,  Katz LC,  Lo DC,  Barney S;
XX
DR   WPI; 2002-025874/03.
DR   P-PSDB; ABB44640.
XX
PT   New protective sequences and their products, useful for diagnosing and
PT   treating diseases involving cell death, including neurological
PT   disorders e.g. stroke and for identifying modulators of expression of
PT   the protective sequences  -
XX
PS   Claim 2; Fig 5; 283pp; English.
XX
CC   The present invention relates to protective sequence proteins
CC   (ABB44624-ABB44830) and their coding sequences (ABA82701-ABA82937).
CC   The sequences, when introduced into a cell either predisposed to undergo
CC   cell death or in the process of undergoing cell death, prevent, delay or
CC   rescue the cell from death, hence, these sequences are named "protective
CC   sequences". The sequences are useful for treating and/or ameliorating
CC   cancer, autoimmune diseases and neurological disorders e.g. stroke.
CC   Further examples of diseases which may be treated by the present
CC   invention are given in the specification.
XX
SQ   Sequence 80 BP; 40 A; 10 C; 15 G; 15 T; 0 other;

  Query Match             2.0%;  Score 29;  DB 24;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 0.035;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db     39 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 67


RESULT 13
ABV08728
ID   ABV08728 standard; cDNA; 223 BP.
XX
AC   ABV08728;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 8719.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 1376; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 223 BP; 76 A; 44 C; 64 G; 38 T; 1 other;

  Query Match             2.0%;  Score 29;  DB 23;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 0.031;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db    188 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 216


RESULT 14
ABL82990/c
ID   ABL82990 standard; cDNA; 311 BP.
XX
AC   ABL82990;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:5968.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US17756.
XX
```

PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
PS Claim 1; SEQ ID 5968; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 311 BP; 96 A; 42 C; 45 G; 128 T; 0 other;

Query Match 2.0%; Score 29; DB 24; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 40 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 12

RESULT 15
ABV13242/c
ID ABV13242 standard; cDNA; 384 BP.
XX
AC ABV13242;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 13233.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 2192; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 384 BP; 136 A; 69 C; 73 G; 106 T; 0 other;

Query Match 2.0%; Score 29; DB 23; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 42 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 14

Search completed: December 14, 2002, 17:02:56
Job time : 290 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 14:01:05 ; Search time 59 Seconds
(without alignments)
7360.242 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	2.1	3300	4	US-09-336-643A-82	Sequence 82, Appl
2	29	2.0	1193	1	US-08-036-555B-134	Sequence 134, App
3	29	2.0	1193	1	US-08-469-569-134	Sequence 134, App
4	29	2.0	1193	1	US-08-249-322A-134	Sequence 134, App
5	29	2.0	1193	1	US-08-469-526A-134	Sequence 134, App
6	29	2.0	1193	2	US-08-734-591A-134	Sequence 134, App
7	29	2.0	1193	2	US-08-469-660-134	Sequence 134, App
8	29	2.0	1193	3	US-08-341-018-3	Sequence 3, Appli
9	29	2.0	1193	3	US-08-470-335-134	Sequence 134, App
10	29	2.0	1193	4	US-08-735-021-134	Sequence 134, App
11	29	2.0	1193	4	US-08-734-664A-134	Sequence 134, App
12	29	2.0	1193	4	US-08-470-339-134	Sequence 134, App
13	29	2.0	1193	4	US-08-467-602-134	Sequence 134, App
14	29	2.0	1193	5	PCT-US94-05083C-130	Sequence 130, App
15	29	2.0	1193	5	PCT-US95-06846A-134	Sequence 134, App
16	29	2.0	4843	3	US-08-986-485-1	Sequence 1, Appli
17	28	2.0	1108	1	US-08-036-555B-135	Sequence 135, App
18	28	2.0	1108	1	US-08-469-569-135	Sequence 135, App
19	28	2.0	1108	1	US-08-249-322A-135	Sequence 135, App
20	28	2.0	1108	1	US-08-469-526A-135	Sequence 135, App
21	28	2.0	1108	2	US-08-734-591A-135	Sequence 135, App
22	28	2.0	1108	2	US-08-469-660-135	Sequence 135, App
23	28	2.0	1108	3	US-08-341-018-5	Sequence 5, Appli
24	28	2.0	1108	3	US-08-470-335-135	Sequence 135, App
25	28	2.0	1108	4	US-08-735-021-135	Sequence 135, App
26	28	2.0	1108	4	US-08-734-664A-135	Sequence 135, App
27	28	2.0	1108	4	US-08-470-339-135	Sequence 135, App
28	28	2.0	1108	4	US-08-467-602-135	Sequence 135, App
29	28	2.0	1108	5	PCT-US94-05083C-131	Sequence 131, App
30	28	2.0	1108	5	PCT-US95-06846A-135	Sequence 135, App
31	28	2.0	3238	4	US-08-123-934A-5	Sequence 5, Appli
32	28	2.0	3238	5	PCT-US94-10080-5	Sequence 5, Appli
33	28	2.0	8201	1	US-08-253-155A-9	Sequence 9, Appli
34	27	1.9	989	2	US-08-874-460-1	Sequence 1, Appli
35	27	1.9	1969	2	US-08-541-033A-7	Sequence 7, Appli
36	27	1.9	1969	2	US-08-828-451-7	Sequence 7, Appli
37	27	1.9	2096	2	US-08-541-033A-19	Sequence 19, Appl
38	27	1.9	2096	2	US-08-828-451-19	Sequence 19, Appl
39	27	1.9	2099	2	US-08-541-033A-3	Sequence 3, Appli
40	27	1.9	2099	2	US-08-828-451-3	Sequence 3, Appli
41	27	1.9	2137	2	US-08-541-033A-18	Sequence 18, Appl
42	27	1.9	2137	2	US-08-828-451-18	Sequence 18, Appl
43	27	1.9	2140	2	US-08-541-033A-1	Sequence 1, Appli
44	27	1.9	2140	2	US-08-828-451-1	Sequence 1, Appli
45	27	1.9	2420	5	PCT-US93-00227-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-336-643A-82
; Sequence 82, Application US/09336643A
; Patent No. 6399761
; GENERAL INFORMATION:
; APPLICANT: Miller, Andrew P.
; APPLICANT: Curran, Mark Edward
; APPLICANT: Hu, Ping
; APPLICANT: Rutter, Marc
; APPLICANT: Wang, Jian-Wang
; TITLE OF INVENTION: No. 6399761el Human Potassium Channels
; FILE REFERENCE: SEQ-15P
; CURRENT APPLICATION NUMBER: US/09/336,643A
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 60/076,687
; PRIOR FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: 60/116,448
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: PCT/US99/03826
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 3300
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1285)
US-09-336-643A-82

Query Match 2.1%; Score 30; DB 4; Length 3300;
Best Local Similarity 100.0%; Pred. No. 0.00092;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 3267

RESULT 2
US-08-036-555B-134
; Sequence 134, Application US/08036555B
; Patent No. 5530109
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use

```
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/036,555B
;      FILING DATE:  24-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APRIL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.K. 91 07566.3
;      FILING DATE:  10-APRIL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsai, Christine H.
;      REGISTRATION NUMBER:  34,266
;      REFERENCE/DOCKET NUMBER:  LUD 5250.4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-036-555B-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 3
US-08-469-569-134
; Sequence 134, Application US/08469569
; Patent No. 5606032
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew; Stroobant, Paul;
;    APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;    APPLICANT:  Chen, Maio Su; Hiles, Ian
;    TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;    TITLE OF INVENTION:  Preparation and Use
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,569
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  24-MAR-1993
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APRIL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.K. 91 07566.3
;      FILING DATE:  10-APRIL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsai, Christine H.
;      REGISTRATION NUMBER:  34,266
;      REFERENCE/DOCKET NUMBER:  LUD 5250.4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-469-569-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 4
US-08-249-322A-134
; Sequence 134, Application US/08249322A
; Patent No. 5716930
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew; Stroobant, Paul;
;    APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;    APPLICANT:  Chen, Maio Su; Hiles, Ian
;    TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;    TITLE OF INVENTION:  Preparation and Use
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
```

```
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/249,322A
;       FILING DATE:  26-MAY-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/036,555
;       FILING DATE:  24-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/965,173
;       FILING DATE:  23-OCT-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/940,389
;       FILING DATE:  03-SEP-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/907,138
;       FILING DATE:  30-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/863,703
;       FILING DATE:  03-APRIL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.K. 91 07566.3
;       FILING DATE:  10-APRIL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Tsai, Christine H.
;       REGISTRATION NUMBER:  34,266
;       REFERENCE/DOCKET NUMBER:  LUD 250.4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  134:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1193
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-249-322A-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 5
US-08-469-526A-134
; Sequence 134, Application US/08469526A
; Patent No. 5792849
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Maio Su
;    APPLICANT:  Hiles, Ian
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,526A
;      FILING DATE:  06 June 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  24-MAR-1993
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  03-JUN-1992
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APRIL-1992
;      APPLICATION NUMBER:  U.K. 91 07566.3
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  04585/00200A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-428-0200
;      TELEFAX:  617-428-7045
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-469-526A-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 6
US-08-734-591A-134
; Sequence 134, Application US/08734591A
; Patent No. 5854220
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Hiles, Ian
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Mario
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM Compatible Pentium
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  WordPerfect (Version 7.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,591A
;      FILING DATE:  22-OCT-1996
;      CLASSIFICATION:  536
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/470,335
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  03-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  UK 91 07566.3
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  04585/00200P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-734-591A-134

  Query Match             2.0%;  Score 29;  DB 2;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 7
US-08-469-660-134
; Sequence 134, Application US/08469660
; Patent No. 5876973
;  GENERAL INFORMATION:
;    APPLICANT:  Gwynne, David I.; Marchionni, Mark;
;    APPLICANT:  McBurney, Robert N.
;    TITLE OF INVENTION:  INHIBITORS OF CELL PROLIFERATION,
;    TITLE OF INVENTION:  THEIR PREPARATION AND USE
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      ZIP:  0211-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,660
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/011,396
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/984,085
;      FILING DATE:  01-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/951,747
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/927,337
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/017004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  200154
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-469-660-134

  Query Match             2.0%;  Score 29;  DB 2;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 8
US-08-341-018-3
; Sequence 3, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
;  APPLICANT: Gywnne, David I.
;  APPLICANT: Mahanthappa, Nagesh K.
;  APPLICANT: Marchionni, Mark A.
;  APPLICANT: Bermingham-McDonogh, Olivia
;  APPLICANT: Goldin, Stanley M.
;  APPLICANT: McBurney, Robert N.
;  TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
;  TITLE OF INVENTION: CELLULAR COMMUNICATION
;  FILE REFERENCE: 04585/041001
;  CURRENT APPLICATION NUMBER: US/08/341,018A
;  CURRENT FILING DATE: 1994-11-17
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-341-018-3

  Query Match             2.0%;  Score 29;  DB 3;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 9
US-08-470-335-134
; Sequence 134, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
```

```
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/00200B
;  CURRENT APPLICATION NUMBER: US/08/470,335F
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  NUMBER OF SEQ ID NOS: 252
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-470-335-134

  Query Match             2.0%;  Score 29;  DB 3;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 10
US-08-735-021-134
; Sequence 134, Application US/08735021B
; Patent No. 6194377
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/00200L
;  CURRENT APPLICATION NUMBER: US/08/735,021B
;  CURRENT FILING DATE: 1996-10-22
;  EARLIER APPLICATION NUMBER: 08/472,065
;  EARLIER FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  EARLIER APPLICATION NUMBER: 07/965,173
;  EARLIER FILING DATE: 1992-10-23
;  EARLIER APPLICATION NUMBER: 07/940,389
;  EARLIER FILING DATE: 1992-09-03
;  EARLIER APPLICATION NUMBER: 07/907,138
;  EARLIER FILING DATE: 1992-06-30
;  EARLIER APPLICATION NUMBER: 07/863,703
;  EARLIER FILING DATE: 1992-04-03
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (8)...(796)
US-08-735-021-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 11
US-08-734-664A-134
; Sequence 134, Application US/08734664A
; Patent No. 6204241
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Mario
;    APPLICANT:  Hiles, Ian
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM Compatible Pentium
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSeq Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,664A
;      FILING DATE:  22-OCT-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/249,322
;      FILING DATE:  26-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  24-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  UK 91 07566.3
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  04585/00200J
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-734-664A-134
```

```
  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 12
US-08-470-339-134
; Sequence 134, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/002008
;  CURRENT APPLICATION NUMBER: US/08/470,339C
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  EARLIER APPLICATION NUMBER: 07/940,389
;  EARLIER FILING DATE: 1992-09-03
;  EARLIER APPLICATION NUMBER: 07/907,138
;  EARLIER FILING DATE: 1992-06-30
;  EARLIER APPLICATION NUMBER: 07/863,703
;  EARLIER FILING DATE: 1992-04-03
;  EARLIER APPLICATION NUMBER: 91 07566.3 GB
;  EARLIER FILING DATE: 1999-04-10
;  NUMBER OF SEQ ID NOS: 226
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-470-339-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 13
US-08-467-602-134
; Sequence 134, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
;  APPLICANT: Sklar, Robert
;  APPLICANT: Marchionni, Mark
;  APPLICANT: Gwynne, David I.
;  TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
;  TITLE OF INVENTION: DISORDERS
;  FILE REFERENCE: 04585/028003
;  CURRENT APPLICATION NUMBER: US/08/467,602C
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/209,204
;  EARLIER FILING DATE: 1994-03-08
;  EARLIER APPLICATION NUMBER: 08/059,022
;  EARLIER FILING DATE: 1993-05-06
;  NUMBER OF SEQ ID NOS: 420
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-467-602-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 14
PCT-US94-05083C-130
; Sequence 130, Application PC/TUS9405083C
;  GENERAL INFORMATION:
;    APPLICANT:  Robert Sklar, Mark Marchionni,
;    APPLICANT:  David I. Gwynne
;    TITLE OF INVENTION:  METHODS FOR ALTERING
;    TITLE OF INVENTION:  MUSCLE CONDITION
;    NUMBER OF SEQUENCES:  185
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360
;      MEDIUM TYPE:  kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05083C
;      FILING DATE:  06-MAY-94
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/209,204
;      FILING DATE:  08-MAR-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/059,022
;      FILING DATE:  06-May-93
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/028WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  130:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US94-05083C-130

  Query Match             2.0%;  Score 29;  DB 5;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 15
PCT-US95-06846A-134
```

```
; Sequence 134, Application PC/TUS9506846A
;  GENERAL INFORMATION:
;     APPLICANT:  Goodearl, Andrew David; Stroobant, Paul;
;     APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;     APPLICANT:  Chen, Maio Su; Hiles, Ian
;     TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;     TITLE OF INVENTION:  Preparation and Use
;     NUMBER OF SEQUENCES:  178
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Felfe & Lynch
;       STREET:  805 Third Avenue
;       CITY:  New York City
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;       COMPUTER:  IBM
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06846A
;       FILING DATE:  25-MAY-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/249,322
;       FILING DATE:  26-MAY-1994
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/036,555
;       FILING DATE:  24-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/965,173
;       FILING DATE:  23-OCT-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/940,389
;       FILING DATE:  03-SEP-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/907,138
;       FILING DATE:  30-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/863,703
;       FILING DATE:  03-APRIL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.K. 91 07566.3
;       FILING DATE:  10-APRIL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, Norman D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5250.5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  134:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1193
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US95-06846A-134

  Query Match             2.0%;  Score 29;  DB 5;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


Search completed: December 14, 2002, 17:04:10
Job time : 78 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 16:09:00 ; Search time 67 Seconds
(without alignments)
8357.803 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 355320 seqs, 197730502 residues

Word size : 0

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1160	81.9	1382	10	US-09-925-300-692	Sequence 692, App
c	2	326	23.0	377	9	US-09-736-457-1096	Sequence 1096, Ap
c	3	326	23.0	377	9	US-09-902-941-1096	Sequence 1096, Ap
c	4	87	6.1	2261	10	US-09-920-300A-1684	Sequence 1684, Ap
c	5	87	6.1	2261	12	US-10-033-528-1684	Sequence 1684, Ap
	6	32	2.3	206	10	US-09-783-590-11058	Sequence 11058, A
c	7	29	2.0	311	10	US-09-867-701-5968	Sequence 5968, Ap
c	8	29	2.0	564	10	US-09-880-107-256	Sequence 256, App
	9	28	2.0	280	10	US-09-960-352-12666	Sequence 12666, A
c	10	28	2.0	319	10	US-09-867-701-2350	Sequence 2350, Ap
	11	28	2.0	941	10	US-09-770-445-359	Sequence 359, App
	12	28	2.0	1505	10	US-09-925-301-142	Sequence 142, App
	13	28	2.0	3238	10	US-09-874-628-5	Sequence 5, Appli
	14	28	2.0	3592	10	US-09-777-745-1	Sequence 1, Appli
	15	28	2.0	3627	10	US-09-777-745-5	Sequence 5, Appli
	16	28	2.0	7878	10	US-09-962-436-562	Sequence 562, App
	17	27	1.9	426	10	US-09-777-564-451	Sequence 451, App
	18	27	1.9	639	9	US-09-950-933A-12	Sequence 12, Appl
	19	27	1.9	915	10	US-09-925-301-244	Sequence 244, App
	20	27	1.9	989	10	US-09-272-162-1	Sequence 1, Appli
	21	27	1.9	1169	10	US-09-804-156-10	Sequence 10, Appl
	22	27	1.9	1334	10	US-09-789-561-30	Sequence 30, Appl
	23	27	1.9	1575	10	US-09-833-381-1159	Sequence 1159, Ap
c	24	27	1.9	1689	9	US-09-989-919-71	Sequence 71, Appl
	25	27	1.9	1969	10	US-09-070-844-7	Sequence 7, Appli
	26	27	1.9	2096	10	US-09-070-844-19	Sequence 19, Appl
	27	27	1.9	2099	10	US-09-070-844-3	Sequence 3, Appli
	28	27	1.9	2137	10	US-09-070-844-18	Sequence 18, Appl
	29	27	1.9	2140	10	US-09-070-844-1	Sequence 1, Appli
	30	27	1.9	2915	9	US-10-042-431-72	Sequence 72, Appl
	31	27	1.9	3441	9	US-09-944-413-6	Sequence 6, Appli
	32	27	1.9	3441	9	US-09-944-403-6	Sequence 6, Appli
	33	27	1.9	3441	9	US-09-944-896-6	Sequence 6, Appli
	34	27	1.9	3441	9	US-09-944-944-6	Sequence 6, Appli
	35	27	1.9	3441	10	US-09-866-028-6	Sequence 6, Appli
	36	27	1.9	3441	10	US-09-944-449-6	Sequence 6, Appli
	37	27	1.9	3441	10	US-09-944-457-6	Sequence 6, Appli
	38	27	1.9	3441	10	US-09-945-587-6	Sequence 6, Appli
	39	27	1.9	3441	10	US-09-945-015-6	Sequence 6, Appli
	40	27	1.9	3441	10	US-09-944-396-6	Sequence 6, Appli
	41	27	1.9	3441	10	US-09-944-097-6	Sequence 6, Appli
	42	27	1.9	3441	10	US-09-944-432-6	Sequence 6, Appli
	43	27	1.9	3441	10	US-09-943-762-6	Sequence 6, Appli
	44	27	1.9	3441	10	US-09-944-654-6	Sequence 6, Appli
	45	27	1.9	3441	10	US-09-943-851A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-925-300-692
; Sequence 692, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
; LENGTH: 1382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-692

Query Match 81.9%; Score 1160; DB 10; Length 1382;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy    230 TTCCTCCTGCGAACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 TTCCTCCTGCGAACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318
```

```
Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGG-CTATGTGTGAAAGCAACTCTAC 617

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 677

Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 737

Qy    770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 797

Qy    830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 857

Qy    890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 917

Qy    950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 977

Qy   1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1037

Qy   1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1097

Qy   1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1157

Qy   1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1217

Qy   1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1277

Qy   1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1337

Qy   1370 GAAAAATATAATAATCATAAA 1390
          ||||||||||||||||||||
Db   1338 GAAAAATATAATAATCATAAA 1358
```

```
RESULT 2
US-09-736-457-1096/c
; Sequence 1096, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang; Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-736-457-1096

  Query Match             23.0%;  Score 326;  DB 9;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-121;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy    189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy    249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy    309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy    369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy    429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1
```

```
RESULT 3
US-09-902-941-1096/c
; Sequence 1096, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
```

```
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-902-941-1096

  Query Match             23.0%;  Score 326;  DB 9;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-121;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy     189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy     249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy     309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy     369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy     429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy     489 GGTTTCCATGAGCGTGG 505
           |||||||||||||||||
Db      17 GGTTTCCATGAGCGTGG 1


RESULT 4
US-09-920-300A-1684/c
; Sequence 1684, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547
;  CURRENT APPLICATION NUMBER: US/09/920,300A
;  CURRENT FILING DATE:  2001-07-31
;  NUMBER OF SEQ ID NOS: 1789
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1684
;   LENGTH: 2261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-920-300A-1684

  Query Match             6.1%;  Score 87;  DB 10;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy    1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db    2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 5
US-10-033-528-1684/c
; Sequence 1684, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547C1
;  CURRENT APPLICATION NUMBER: US/10/033,528
;  CURRENT FILING DATE:  2001-12-26
;  NUMBER OF SEQ ID NOS: 1896
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1684
;   LENGTH: 2261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-033-528-1684

  Query Match             6.1%;  Score 87;  DB 12;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy    1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db    2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 6
US-09-783-590-11058
; Sequence 11058, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11058
;   LENGTH: 206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: misc feature
;   LOCATION: (5)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (47)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (52)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (60)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (76)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (149)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (164)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (172)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (187)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (190)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (198)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (204)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-11058

  Query Match             2.3%;  Score 32;  DB 10;  Length 206;
  Best Local Similarity   100.0%;  Pred. No. 0.00065;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     79 GTAGGACCTACTCCCTATCCCGTCGGCCGCGG 110
          ||||||||||||||||||||||||||||||||
Db      8 GTAGGACCTACTCCCTATCCCGTCGGCCGCGG 39


RESULT 7
US-09-867-701-5968/c
; Sequence 5968, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5968
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-5968

  Query Match             2.0%;  Score 29;  DB 10;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 0.0099;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db     40 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 12


RESULT 8
US-09-880-107-256/c
; Sequence 256, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 256
;   LENGTH: 564
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA127741
US-09-880-107-256

  Query Match             2.0%;  Score 29;  DB 10;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 0.0094;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db     43 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 15


RESULT 9
US-09-960-352-12666
; Sequence 12666, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12666
;   LENGTH: 280
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 54-LIB3058-028-Q1-K1-F6
US-09-960-352-12666

  Query Match             2.0%;  Score 28;  DB 10;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db    200 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 227


RESULT 10
US-09-867-701-2350/c
```

; Sequence 2350, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2350
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(319)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2350

Query Match 2.0%; Score 28; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||
Db 43 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 16

RESULT 11
US-09-770-445-359
; Sequence 359, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 941
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(941)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-445-359

Query Match 2.0%; Score 28; DB 10; Length 941;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||
Db 912 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 939

RESULT 12
US-09-925-301-142
; Sequence 142, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1493)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1499)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1500)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-142

Query Match 2.0%; Score 28; DB 10; Length 1505;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||
Db 1432 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1459

RESULT 13
US-09-874-628-5
; Sequence 5, Application US/09874628
; Patent No. US20020137133A1
; GENERAL INFORMATION:
; APPLICANT: WOZNEY, John
; CELESTE, Anthony J.
; THIES, R. Scott
; YAMAJI, No. US20020137133A1oru
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/874,628
;             FILING DATE: 05-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/123,934
;             FILING DATE: 17-SEP-1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: LAZAR, Steven R
;             REGISTRATION NUMBER: 32,618
;             REFERENCE/DOCKET NUMBER: 5203
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617 876 1170
;             TELEFAX: 617 876 5851
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3238 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        IMMEDIATE SOURCE:
;             CLONE: CFK1-10a
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  474..2000
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-874-628-5

  Query Match             2.0%;  Score 28;  DB 10;  Length 3238;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3155 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3182


RESULT 14
US-09-777-745-1
; Sequence 1, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
;  APPLICANT:  CALMELS, THIERRY PAUL GERARD
;  APPLICANT:  SOUCHET, MICHEL LOUIS
;  APPLICANT:  LEGER, ISABELLE MARIE
;  APPLICANT:  TOURTELIER, LAURENCE NATHALIE PATRICIA
;  APPLICANT:  BRIL, ANTOINE MICHEL ALAIN
;  APPLICANT:  DOE, TRUDY RACHEL
;  TITLE OF INVENTION:  NOVEL COMPOUNDS
;  FILE REFERENCE:  GH-30321-C1
;  CURRENT APPLICATION NUMBER: US/09/777,745
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER:  EP 98400014.1
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  EP 98400015.8
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  UK 9710910.2
;  PRIOR FILING DATE:  1997-05-27
;  PRIOR APPLICATION NUMBER:  09/082,271
;  PRIOR FILING DATE:  1998-05-20
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 3592
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-777-745-1

  Query Match             2.0%;  Score 28;  DB 10;  Length 3592;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3559 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3586


RESULT 15
US-09-777-745-5
; Sequence 5, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
;  APPLICANT:  CALMELS, THIERRY PAUL GERARD
;  APPLICANT:  SOUCHET, MICHEL LOUIS
;  APPLICANT:  LEGER, ISABELLE MARIE
;  APPLICANT:  TOURTELIER, LAURENCE NATHALIE PATRICIA
;  APPLICANT:  BRIL, ANTOINE MICHEL ALAIN
;  APPLICANT:  DOE, TRUDY RACHEL
;  TITLE OF INVENTION:  NOVEL COMPOUNDS
;  FILE REFERENCE:  GH-30321-C1
;  CURRENT APPLICATION NUMBER: US/09/777,745
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER:  EP 98400014.1
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  EP 98400015.8
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  UK 9710910.2
;  PRIOR FILING DATE:  1997-05-27
;  PRIOR APPLICATION NUMBER:  09/082,271
;  PRIOR FILING DATE:  1998-05-20
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3627
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-777-745-5

  Query Match             2.0%;  Score 28;  DB 10;  Length 3627;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3594 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3621


Search completed: December 14, 2002, 17:56:17
Job time : 102 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 16:58:17 ; Search time 2016 Seconds
(without alignments)
11375.401 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	844	59.6	844	9	AL529152	AL529152 AL529152
	2	813	57.4	890	9	AL520768	AL520768 AL520768
	3	798	56.4	943	14	BM804704	BM804704 AGENCOURT
c	4	691	48.8	960	9	AL522373	AL522373 AL522373
	5	677	47.8	782	14	BQ441752	BQ441752 AGENCOURT
	6	675	47.7	1048	14	BQ059007	BQ059007 AGENCOURT
c	7	660	46.6	677	14	BQ581760	BQ581760 il12g06.x
c	8	645	45.6	659	13	BM666784	BM666784 UI-E-CL1-
c	9	625	44.1	625	13	BI711300	BI711300 id96f12.x
c	10	610	43.1	729	14	BM975958	BM975958 UI-CF-EN1
	11	606	42.8	613	14	BM690873	BM690873 UI-E-CI1-
	12	604	42.7	846	13	BI769604	BI769604 603054985
c	13	589	41.6	750	14	BM979227	BM979227 UI-CF-DU1
c	14	585	41.3	992	9	AL520767	AL520767 AL520767
c	15	584	41.2	586	9	AI797259	AI797259 we86e02.x
	16	577	40.7	600	14	BM726885	BM726885 UI-E-EJ0-
	17	573	40.5	1150	14	BM925975	BM925975 AGENCOURT
	18	572	40.4	579	13	BI711712	BI711712 id96f12.y
c	19	571	40.3	736	9	AI668782	AI668782 wc14b12.x
	20	563	39.8	624	14	BQ582098	BQ582098 il12g06.y
c	21	557	39.3	559	13	BM141822	BM141822 if26h04.x
c	22	548	38.7	550	13	BM141736	BM141736 if25d10.x
c	23	547	38.6	766	14	BM679998	BM679998 UI-E-EO1-
	24	545	38.5	793	12	BG436916	BG436916 602488424
	25	543	38.3	746	13	BI752623	BI752623 603028471
	26	539	38.1	938	14	BQ073413	BQ073413 AGENCOURT
	27	534	37.7	764	12	BG762506	BG762506 602733954
c	28	524	37.0	550	9	AI017543	AI017543 ou35c05.x
c	29	510	36.0	710	9	AI797380	AI797380 we87d12.x
	30	510	36.0	781	12	BG532265	BG532265 602561355
	31	510	36.0	928	13	BI822844	BI822844 603040166
	32	505	35.7	537	13	BM142010	BM142010 if25d10.y
	33	503	35.5	552	13	BM142111	BM142111 if26h04.y
c	34	487	34.4	560	10	AW182514	AW182514 xj81b03.x
	35	468	33.1	576	10	AW956706	AW956706 EST368776
	36	468	33.1	657	13	BI916460	BI916460 603182527
c	37	461	32.6	461	9	AA907376	AA907376 oj70h11.s
	38	459	32.4	946	13	BI757823	BI757823 603030404
	39	457	32.3	790	12	BG703176	BG703176 602686082
	40	452	31.9	813	13	BI755493	BI755493 603027349
c	41	451	31.9	633	14	BQ002524	BQ002524 UI-H-EI1-
	42	448	31.6	791	12	BG776560	BG776560 602663693
c	43	446	31.5	579	9	AI559188	AI559188 tq42g05.x
c	44	444	31.4	749	10	AW073366	AW073366 xa39c01.x
c	45	434	30.6	482	9	AI085503	AI085503 qf20c07.x

ALIGNMENTS

RESULT 1
AL529152
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5 prime, mRNA sequence.
ACCESSION AL529152
VERSION AL529152.1 GI:12792645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YM06"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

```
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      182 a    238 c    235 g    189 t
ORIGIN

  Query Match           59.6%;  Score 844;  DB 9;  Length 844;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  844;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     300 AAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTG 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTG 60

Qy     360 CCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCA 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCA 120

Qy     420 CTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 180

Qy     480 TCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAAT 539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAAT 240

Qy     540 TGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTG 599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTG 300

Qy     600 TGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGG 659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGG 360

Qy     660 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTAT 719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTAT 420

Qy     720 GACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACC 779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACC 480

Qy     780 TATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCC 839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCC 540

Qy     840 CGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAG 899
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAG 600

Qy     900 AAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGC 959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGC 660

Qy     960 TGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTG 1019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTG 720

Qy    1020 CGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCAC 1079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCAC 780

Qy    1080 CCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGC 1139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGC 840

Qy    1140 CACT 1143
           ||||
Db     841 CACT 844


RESULT 2
AL520768
LOCUS       AL520768                890 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
            prime, mRNA sequence.
ACCESSION   AL520768
VERSION     AL520768.1  GI:12784261
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostcmi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      191 a    251 c    241 g    206 t      1 others
ORIGIN

  Query Match           57.4%;  Score 813;  DB 9;  Length 890;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  813;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     118 TCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTC 177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 TCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTC 129

Qy     178 AACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCT 237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 AACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCT 189

Qy     238 GCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACT 297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 GCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACT 249

Qy     298 TTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 357
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     250 TTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 309

Qy     358 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAAC 417
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAAC 369

Qy     418 CACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGC 477
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    370 CACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGC 429

Qy    478 TGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACA 537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACA 489

Qy    538 ATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATC 597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 ATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATC 549

Qy    598 TGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 TGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 609

Qy    658 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCT 717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCT 669

Qy    718 ATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTA 777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 ATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTA 729

Qy    778 CCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGG 837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 CCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGG 789

Qy    838 CCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATG 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 CCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATG 849

Qy    898 AGAAGAGAGCAAGTTATGGCCACAGCATGGTGG 930
          ||||||||||||||||||||||||||||||||
Db    850 AGAAGAGAGCAAGTTATGGCCACAGCATGGTGG 882
```

```
RESULT 3
BM804704
LOCUS       BM804704                 943 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6494868 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725375
            5', mRNA sequence.
ACCESSION   BM804704
VERSION     BM804704.1  GI:19121527
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 943)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12716  row: c  column: 08
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1. .943
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5725375"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
BASE COUNT      198 a    270 c    260 g    215 t
ORIGIN

  Query Match             56.4%;  Score 798;  DB 14;  Length 943;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  798;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 113

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 173

Qy    230 TTCCTCCTGCGAACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 233

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 293

Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    294 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 353

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 413

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    414 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 473

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 533

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 593

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 653

Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 713

Qy    770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    714 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 773

Qy    830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    774 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 833

Qy    890 CCACCATGAGAAGAGAGC 907
          ||||||||||||||||||
Db    834 CCACCATGAGAAGAGAGC 851
```

```
RESULT 4
AL522373/c
LOCUS       AL522373                 960 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3
```

```
            prime, mRNA sequence.
ACCESSION   AL522373
VERSION     AL522373.1  GI:12785866
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB008YM02"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      225 a    259 c    255 g    217 t      4 others
ORIGIN

  Query Match           48.8%;  Score 691;  DB 9;  Length 960;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches  911;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     438 GAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCAT 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     922 GAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCAT 863

Qy     498 GAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     862 GAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 803

Qy     558 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCA 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     802 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCA 743

Qy     618 GGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCT 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     742 GGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCT 683

Qy     678 GTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAA 737
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     682 GTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGC-GTTCCCTGAA 624

Qy     738 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGA 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     623 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGA 564

Qy     798 TCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     563 TCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 504

Qy     858 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGC 917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     503 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGC 444

Qy     918 CACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGC 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     443 CACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGC 384

Qy     978 CTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTG 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     383 CTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTG 324

Qy    1038 TTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTG 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     323 TTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTG 264

Qy    1098 ACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTC 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     263 ACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTC 204

Qy    1158 ATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCT 1217
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     203 ATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCMCTTGGAGAACCTTGACTCTCT 144

Qy    1218 TGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTC 1277
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     143 TGATGGAACACAGATGGGCTGCTTGGGAAAGAMACTTTCACCTGAGCTTCACCTGAGGTC 84

Qy    1278 AGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGA 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83 AGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGA 24

Qy    1338 AGTTCTGCTGAGGG 1351
           ||||||||||||||
Db      23 AGTTCTGCTGAGGG 10


RESULT 5
BQ441752
LOCUS       BQ441752                782 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7761426 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6019908
            5', mRNA sequence.
ACCESSION   BQ441752
VERSION     BQ441752.1  GI:21180828
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13223  row: c  column: 13
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .782
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6019908"
                     /clone_lib="NIH_MGC_70"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
```

```
BASE COUNT      185 a    209 c    200 g    186 t      2 others
ORIGIN

  Query Match             47.8%;  Score 677;  DB 14;  Length 782;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  677;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     696 AAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCT 755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCT 60

Qy     756 CAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCC 815
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCC 120

Qy     816 CACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCA 875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCA 180

Qy     876 GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGAC 935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGAC 240

Qy     936 CCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATA 995
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATA 300

Qy     996 GACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCT 1055
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCT 360

Qy    1056 GACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC 1115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC 420

Qy    1116 CTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGG 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGG 480

Qy    1176 ATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 1235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 540

Qy    1236 CTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAA 1295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAA 600

Qy    1296 GGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGA 1355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGA 660

Qy    1356 GCAGCACTGGCATTGAA 1372
           |||||||||||||||||
Db     661 GCAGCACTGGCATTGAA 677


RESULT 6
BQ059007
LOCUS       BQ059007               1048 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6808192 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5814981
            5', mRNA sequence.
ACCESSION   BQ059007
VERSION     BQ059007.1  GI:19818347
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1048)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2069  row: h  column: 22
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5814981"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      216 a    304 c    289 g    239 t
ORIGIN

  Query Match             47.7%;  Score 675;  DB 14;  Length 1048;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  675;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 124

Qy     170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 184

Qy     230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 244

Qy     290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 304

Qy     350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 364

Qy     410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 424

Qy     470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     425 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 484

Qy     530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 544

Qy     590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 604

Qy     650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     605 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 664
```

```
Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    665 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 724

Qy    770 GATACTTACCTATCC 784
          ||||||||||||||||
Db    725 GATACTTACCTATCC 739


RESULT 7
BQ581760/c
LOCUS       BQ581760                677 bp    mRNA    linear   EST 20-JUN-2002
DEFINITION  il12g06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029771
            3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1. ;, mRNA
            sequence.
ACCESSION   BQ581760
VERSION     BQ581760.1  GI:21494649
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6029771"
                     /clone_lib="Human insulinoma"
                     /tissue_type="insulinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
                     (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
                     pBluescript SK- by Dr. H. Inoue following the Washington
                     University protocol
                     (http://genome.wustl.edu/est/lambda_protocol.shtml).
                     Please contact Hiroshi Inoue, MD/PhD for further
                     information on this library (Metabolism Division, Permutt
                     Laboratory, Washington University School of Medicine, Box
                     8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
                     is a Washington University Pancreas EST project library. "
BASE COUNT      158 a    170 c    179 g    170 t
ORIGIN

  Query Match           46.6%;  Score 660;  DB 14;  Length 677;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  660;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    732 CCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 CCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 618

Qy    792 TTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAA 851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 TTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAA 558

Qy    852 ACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGT 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGT 498

Qy    912 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGG 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGG 438

Qy    972 CCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTG 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 CCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTG 378

Qy   1032 CCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAA 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAA 318

Qy   1092 GACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTA 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db.   317 GACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTA 258

Qy   1152 GTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGA 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 GTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGA 198

Qy   1212 CTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCT 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 CTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCT 138

Qy   1272 GAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGA 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 GAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGA 78

Qy   1332 AACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 18


RESULT 8
BM666784/c
LOCUS       BM666784                659 bp    mRNA    linear   EST 27-FEB-2002
DEFINITION  UI-E-CL1-aez-a-24-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
            UI-E-CL1-aez-a-24-0-UI 3', mRNA sequence.
ACCESSION   BM666784
VERSION     BM666784.1  GI:18974615
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
```

```
          Clone Distribution: Researchers may obtain clones from Research
          Genetics (www.resgen.com).
          The following repetitive elements were found in this cDNA
          sequence: 1-35, >AT_rich#Low_complexity
          Seq primer: M13 Forward
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL1-aez-a-24-0-UI"
                     /clone_lib="UI-E-CL1"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CL1 is a normalized cDNA library containing the
                     following tissue(s): retina. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CCGCG. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI).
                     TAG_LIB=UI-E-CL1
                     TAG_TISSUE=human retina
                     TAG_SEQ=CCGCG"
BASE COUNT      153 a    168 c    173 g    165 t
ORIGIN

  Query Match             45.6%;  Score 645;  DB 13;  Length 659;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  645;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     747 GCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     659 GCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 600

Qy     807 GGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTA 866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     599 GGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTA 540

Qy     867 GTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATG 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 GTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATG 480

Qy     927 GTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTT 986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 GTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTT 420

Qy     987 GCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCAC 1046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 GCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCAC 360

Qy    1047 CGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTG 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 CGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTG 300

Qy    1107 AGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTT 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 AGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTT 240

Qy    1167 GGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAAC 1226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 GGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAAC 180

Qy    1227 ACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAG 1286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 ACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAG 120

Qy    1287 TTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCT 1346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119 TTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCT 60

Qy    1347 GAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 1391
           |||||||||||||||||||||||||||||||||||||||||||||
Db      59 GAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 15


RESULT 9
BI711300/c
LOCUS       BI711300                625 bp    mRNA    linear   EST 11-MAR-2002
DEFINITION  id96f12.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023654
            3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1. ;, mRNA
            sequence.
ACCESSION   BI711300
VERSION     BI711300.1  GI:15686995
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5023654"
                     /clone_lib="Human insulinoma"
                     /tissue_type="insulinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
                     (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
                     pBluescript SK- by Dr. H. Inoue following the Washington
                     University protocol
                     (http://genome.wustl.edu/est/lambda_protocol.shtml).
                     Please contact Hiroshi Inoue, MD/PhD for further
                     information on this library (Metabolism Division, Permutt
                     Laboratory, Washington University School of Medicine, Box
                     8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
                     is a Washington University Pancreas EST project library. "
BASE COUNT      139 a    183 c    161 g    142 t
ORIGIN

  Query Match             44.1%;  Score 625;  DB 13;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  625;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     333 AGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCT 392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     625 AGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCT 566

Qy     393 GCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAG 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     565 GCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAG 506

Qy     453 CTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGAC 512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     505 CTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGAC 446

Qy     513 TGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTA 572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     445 TGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTA 386

Qy     573 GTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 632
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     385 GTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 326

Qy     633 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCA 692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCA 266

Qy     693 GGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTG 752
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 GGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTG 206

Qy     753 GCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCA 812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 GCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCA 146

Qy     813 GCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCA 872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 GCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCA 86

Qy     873 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTA 932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTA 26

Qy     933 GACCCCTGGGGAACAGTGGTGGCCC 957
           |||||||||||||||||||||||||
Db      25 GACCCCTGGGGAACAGTGGTGGCCC 1
```

```
RESULT 10
BM975958/c
LOCUS       BM975958                729 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-EN1-acw-1-14-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
            UI-CF-EN1-acw-1-14-0-UI 3', mRNA sequence.
ACCESSION   BM975958
VERSION     BM975958.1  GI:19593549
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 1-35, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EN1-acw-1-14-0-UI"
                     /clone_lib="UI-CF-EN1"
                     /tissue_type="Primary Lung Cystic Fibrosis Epithelial
                     Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EN1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Cystic Fibrosis
                     Epithelial Cells. The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was ligated to an EcoR I adaptor, digested with Not
                     I, and cloned directionally into pT7T3-Pac vector. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CTGCTCAGGT.
                     TAG_LIB=UI-CF-EN1
                     TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
                     6hr to LPS 24h
                     TAG_SEQ=CTGCTCAGGT"
BASE COUNT      171 a    183 c    195 g    179 t      1 others
ORIGIN
```

```
  Query Match             43.1%;  Score 610;  DB 14;  Length 729;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  660;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     731 CCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGC 790
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     675 CCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGC 616

Qy     791 TTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGA 850
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     615 TTNTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGA 556

Qy     851 AACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     555 AACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 496

Qy     911 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGG 970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     495 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGG 436

Qy     971 GCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCT 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     435 GCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCT 376

Qy    1031 GCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTA 1090
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     375 GCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTA 316

Qy    1091 AGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCT 1150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 AGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCT 256

Qy    1151 AGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTG 1210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    255 AGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTG 196

Qy   1211 ACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACC 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 ACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACC 136

Qy   1271 TGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGG 1330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 TGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGG 76

Qy   1331 AAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAA 1390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 AAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAA 16

Qy   1391 G 1391
          |
Db     15 G 15
```

```
RESULT 11
BM690873
LOCUS       BM690873                613 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CI1-aaz-a-06-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone
            UI-E-CI1-aaz-a-06-0-UI 5', mRNA sequence.
ACCESSION   BM690873
VERSION     BM690873.1  GI:19004131
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CI1-aaz-a-06-0-UI"
                     /clone_lib="UI-E-CI1"
                     /tissue_type="RPE and Choroid"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CI1 is a normalized cDNA library containing the
                     following tissue(s): RPE and Choroid. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is ACCTA.
                     This library was created for the program, Gene Discovery
                     in the Visual System, supported by National Eye Institute
                     (NEI)."
BASE COUNT      128 a    184 c    164 g    137 t
ORIGIN
```

```
  Query Match           42.8%;  Score 606;  DB 14;  Length 613;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  606;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    645 TCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGT 704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8 TCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGT 67

Qy    705 CTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGA 764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 CTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGA 127

Qy    765 GCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAG 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAG 187

Qy    825 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 247

Qy    885 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 307

Qy    945 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 1004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 367

Qy   1005 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 1064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 427

Qy   1065 GGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTC 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    428 GGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTC 487

Qy   1125 CCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCA 1184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 CCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCA 547

Qy   1185 CAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGG 1244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 CAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGG 607

Qy   1245 AAAGAA 1250
          ||||||
Db    608 AAAGAA 613
```

```
RESULT 12
BI769604
LOCUS       BI769604                846 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603054985F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
            mRNA sequence.
ACCESSION   BI769604
VERSION     BI769604.1  GI:15761182
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11512  row: j  column: 22
            High quality sequence stop: 838.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5204469"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      179 a    240 c    226 g    201 t
ORIGIN

  Query Match             42.7%;  Score 604;  DB 13;  Length 846;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  604;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 120

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 180

Qy    230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 240

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 300

Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 360

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 420

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 480

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 540

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 600

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 660

Qy    710 TGTC 713
          ||||
Db    661 TGTC 664


RESULT 13
BM979227/c
LOCUS       BM979227                750 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-DU1-adl-n-04-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
            UI-CF-DU1-adl-n-04-0-UI 3', mRNA sequence.
ACCESSION   BM979227
VERSION     BM979227.1  GI:19599461
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 1-41, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-DU1-adl-n-04-0-UI"
                     /clone_lib="UI-CF-DU1"
                     /tissue_type="Primary Lung Epithelial Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-DU1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Epithelial Cells The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site.  Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GGCTGTAGGC.
                     TAG_LIB=UI-CF-DU1
                     TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
                     TAG_SEQ=GGCTGTAGGC"
BASE COUNT      185 a    171 c    193 g    200 t      1 others
ORIGIN

  Query Match             41.6%;  Score 589;  DB 14;  Length 750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  589;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     589 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 530

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     529 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 470

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     469 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 410

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     409 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 350

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     349 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 290

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     289 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 230

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     229 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 170

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 110

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     109 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 50

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAA 1411
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAA 1
```

RESULT 14
AL520767/c
LOCUS AL520767 992 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 3 prime, mRNA sequence.
ACCESSION AL520767
VERSION AL520767.1 GI:12784260
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 992)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB002YD05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 229 a 267 c 268 g 222 t 6 others
ORIGIN

Query Match 41.3%; Score 585; DB 9; Length 992;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 805; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy     443 ATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCG 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 ATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCG 870

Qy     503 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAA 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAA 810

Qy     563 AGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 AGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 750

Qy     623 GGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 GGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 690

Qy     683 CACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTC 742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 CACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTC 630

Qy     743 TCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCAT 802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 TCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCAT 570

Qy     803 TACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569 TACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 510

Qy     863 TGTAGTGGC-AGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 921
           ||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 TGTAGTGCCMCGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 450

Qy     922 GCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCT 981
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 GCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCT 390

Qy     982 GCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCC 1041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 GCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCC 330

Qy    1042 AGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTT 1101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 AGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTT 270

Qy    1102 CTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGT 1161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 CTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGT 210

Qy    1162 GACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGAT 1221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 GACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGAT 150

Qy    1222 GGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||
Db     149 GGAACACAGATGGGCTGCTTGGGAAAGA 122
```

RESULT 15
AI797259/c

```
LOCUS       AI797259                586 bp    mRNA    linear   EST 18-DEC-1999
DEFINITION  we86e02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2347994 3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1.
            ;, mRNA sequence.
ACCESSION   AI797259
VERSION     AI797259.1  GI:5362731
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1500   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2347994"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      137 a    155 c    158 g    136 t
ORIGIN
```

Query Match 41.2%; Score 584; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    810 CCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTG 869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 CCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTG 527

Qy    870 GCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTG 929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 GCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTG 467

Qy    930 GTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCC 989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 GTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCC 407

Qy    990 CGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGC 1049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 CGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGC 347

Qy   1050 AGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGT 1109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 AGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGT 287

Qy   1110 TTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGA 1169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 TTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGA 227
```

```
Qy   1170 GGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACA 1229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     226 GGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACA 167

Qy   1230 GATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTT 1289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     166 GATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTT 107

Qy   1290 CAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAG 1349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     106 CAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAG 47

Qy   1350 GGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||||||||||||||||||||
Db      46 GGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 3
```

Search completed: December 14, 2002, 18:30:06
Job time : 2043 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:37:34 ; Search time 2740 Seconds
(without alignments)
15039.990 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1307	92.3	1385	9	AF069987	AF069987 Homo sapi
2		571	40.3	4079	9	AF069984	AF069984 Homo sapi
3	c	571	40.3	167863	2	AL590651	AL590651 Homo sapi
4		571	40.3	200822	9	AL591806	AL591806 Human DNA
5	c	326	23.0	377	6	AX368386	AX368386 Sequence
6	c	173	12.2	441	11	G13436	G13436 human STS W
7	c	87	6.1	2261	6	AX397469	AX397469 Sequence
8	c	87	6.1	2261	9	AK001497	AK001497 Homo sapi
9		47	3.3	847	11	G72919	G72919 MARC 2849-2
10		33	2.3	1338	10	AF069988	AF069988 Mus muscu
11		33	2.3	1365	10	BC021634	BC021634 Mus muscu
12		33	2.3	4481	10	AF069985	AF069985 Mus muscu
13		33	2.3	181583	2	AC105589	AC105589 Rattus no
14		33	2.3	211772	10	AC084821	AC084821 Mus muscu
15	c	33	2.3	215043	10	AC087229	AC087229 Mus muscu
16	c	32	2.3	115667	2	AC125857	AC125857 Rattus no
17		31	2.2	144986	2	AC108571	AC108571 Rattus no
18		30	2.1	1587	9	BC017413	BC017413 Homo sapi
19		30	2.1	2421	10	BC016425	BC016425 Mus muscu
20		30	2.1	3300	6	AR212407	AR212407 Sequence
21		30	2.1	4113	10	AF230376	AF230376 Meriones
22	c	30	2.1	44645	9	AC108058	AC108058 Homo sapi
23		30	2.1	60773	2	AC069576	AC069576 Homo sapi
24	c	30	2.1	252059	2	AC102081	AC102081 Mus muscu
25		29	2.0	345	10	RNRIPRL38	X57007 Rat mRNA fo
26		29	2.0	549	9	AF112344	AF112344 Homo sapi
27	c	29	2.0	564	6	AX407609	AX407609 Sequence
28		29	2.0	876	5	AF175983	AF175983 Rana sylv
29		29	2.0	1049	8	AF071889	AF071889 Prunus ar
30		29	2.0	1075	9	BC014385	BC014385 Homo sapi
31		29	2.0	1134	10	AF058799	AF058799 Mus muscu
32		29	2.0	1193	6	A29423	A29423 putative bo
33		29	2.0	1193	6	AR022483	AR022483 Sequence
34		29	2.0	1193	6	AR068961	AR068961 Sequence
35		29	2.0	1193	6	AR103330	AR103330 Sequence
36		29	2.0	1193	6	AR134988	AR134988 Sequence
37		29	2.0	1193	6	AR141817	AR141817 Sequence
38		29	2.0	1193	6	AR143334	AR143334 Sequence
39		29	2.0	1193	6	AR151943	AR151943 Sequence
40		29	2.0	1193	6	I36423	I36423 Sequence 13
41		29	2.0	1193	6	I88123	I88123 Sequence 13
42		29	2.0	1211	5	AY065841	AY065841 Danio rer
43		29	2.0	1267	10	BC004064	BC004064 Mus muscu
44		29	2.0	1488	9	HSU79258	U79258 Human clone
45		29	2.0	1667	3	AY047568	AY047568 Drosophil

ALIGNMENTS

RESULT 1
AF069987
LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.
ACCESSION AF069987
VERSION AF069987.1 GI:3228665
KEYWORDS .
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1385)
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

```
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            1. .1385
                     /gene="NIT1"
     CDS             77. .1060
                     /gene="NIT1"
                     /note="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC39907.1"
                     /db_xref="GI:3228666"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
BASE COUNT      322 a    382 c    357 g    324 t
ORIGIN

  Query Match             92.3%;  Score 1307;  DB 9;  Length 1385;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1307;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy     170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy     230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy     290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318

Qy     350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy     410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy     470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy     530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy     590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 618

Qy     650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     619 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 678

Qy     710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     679 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 738

Qy     770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     739 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 798

Qy     830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     799 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 858

Qy     890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     859 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 918

Qy     950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     919 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 978

Qy    1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     979 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1038

Qy    1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1039 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1098

Qy    1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1099 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1158

Qy    1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1159 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1218

Qy    1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1219 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1278

Qy    1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1279 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1338

Qy    1370 GAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1339 GAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1385


RESULT 2
AF069984
LOCUS       AF069984                4079 bp    DNA     linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069984
VERSION     AF069984.1  GI:3242977
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
```

```
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            635. .3569
                     /gene="NIT1"
     mRNA            join(635. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3569)
                     /gene="NIT1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            635. .680
                     /gene="NIT1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(679. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3267)
                     /gene="NIT1"
                     /note="Nit1; alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC39901.1"
                     /db_xref="GI:3242978"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
     exon            681. .711
                     /gene="NIT1"
                     /note="1A; alternatively spliced"
     exon            712. .741
                     /gene="NIT1"
                     /note="1B; alternatively spliced"
     exon            940. .1053
                     /gene="NIT1"
                     /note="1C; alternatively spliced"
     exon            1288. .1383
                     /gene="NIT1"
                     /number=2
     exon            1636. .1890
                     /gene="NIT1"
                     /number=3
     exon            2011. .2114
                     /gene="NIT1"
                     /number=4
     exon            2321. .2454
                     /gene="NIT1"
                     /number=5
     exon            2609. .2734
                     /gene="NIT1"
                     /number=6
     exon            3001. .3569
                     /gene="NIT1"
                     /number=7
BASE COUNT      940 a   1082 c   1043 g   1012 t      2 others
ORIGIN

  Query Match             40.3%;  Score 571;  DB 9;  Length 4079;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2999 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 3058

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3059 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 3118

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3119 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 3178

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3179 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 3238

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3239 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 3298

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3299 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 3358

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3359 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 3418

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3419 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 3478

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3479 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 3538

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||||
Db    3539 TGGCATTGAAAAATATAATAATCATAAAGTC 3569


RESULT 3
AL590651/c
LOCUS       AL590651              167863 bp    DNA     linear   HTG 18-AUG-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL590651
VERSION     AL590651.4  GI:13990236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167863)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13751001.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA137A12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
```

```
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 164729 bases at least Q40
          Consensus quality: 165486 bases at least Q30
          Consensus quality: 166169 bases at least Q20
          Insert size: 167063; sum-of-contigs
          Insert size: 171878; 3.4% error; agarose-fp
          Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
          coverage: 8.11x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     9373: contig of 9373 bp in length
          *     9374 9473: gap of      100 bp
          *     9474    29891: contig of 20418 bp in length
          *    29892 29991: gap of      100 bp
          *    29992    36239: contig of 6248 bp in length
          *    36240 36339: gap of      100 bp
          *    36340   103784: contig of 67445 bp in length
          *   103785 103884: gap of      100 bp
          *   103885   114878: contig of 10994 bp in length
          *   114879 114978: gap of      100 bp
          *   114979   126347: contig of 11369 bp in length
          *   126348 126447: gap of      100 bp
          *   126448   137764: contig of 11317 bp in length
          *   137765 137864: gap of      100 bp
          *   137865   163160: contig of 25296 bp in length
          *   163161 163260: gap of      100 bp
          *   163261   167863: contig of 4603 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-137A12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .9373
                     /note="assembly_fragment:02360
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9474. .29891
                     /note="assembly_fragment:01854
                     fragment_chain:1"
     misc_feature    29992. .36239
                     /note="assembly_fragment:00805
                     fragment_chain:1"
     misc_feature    36340. .103784
                     /note="assembly_fragment:01467
                     fragment_chain:1"
     misc_feature    103885. .114878
                     /note="assembly_fragment:00991
                     fragment_chain:2"
     misc_feature    114979. .126347
                     /note="assembly_fragment:01503
                     fragment_chain:2"
     misc_feature    126448. .137764
                     /note="assembly_fragment:01297
                     fragment_chain:2"
     misc_feature    137865. .163160
                     /note="assembly_fragment:00117
                     fragment_chain:2"
     misc_feature    163261. .167863
                     /note="assembly_fragment:01267
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    43927 a  39935 c  40002 g  43197 t    802 others
```

```
ORIGIN

  Query Match             40.3%;  Score 571;  DB 2;  Length 167863;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149682 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 149623

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149622 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 149563

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149562 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 149503

Qy    1003 ACTATCTGCGACAGTTGCGCCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149502 ACTATCTGCGACAGTTGCGCCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 149443

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149442 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 149383

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149382 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 149323

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149322 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 149263

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149262 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 149203

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149202 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 149143

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||
Db 149142 TGGCATTGAAAAATATAATAATCATAAAGTC 149112


RESULT 4
AL591806
LOCUS       AL591806              200822 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
            sequence.
ACCESSION   AL591806
VERSION     AL591806.16  GI:18476709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200822)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17902927.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
```

```
           regions were either double-stranded or sequenced with an alternate
           chemistry or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by at least
           one plasmid subclone or more than one M13 subclone; and the
           assembly was confirmed by restriction digest. The following
           abbreviations are used to associate primary accession numbers given
           in the feature table with their source databases: Em:, EMBL; Sw:,
           SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
           database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
           Group.  Further information can be found at
           http://www.sanger.ac.uk/HGP/Chr1
           RP11-544M22 is from the library RPCI-11.2 constructed by the group
           of Pieter de Jong. For further details see
           http://www.chori.org/bacpac/home.htm
           VECTOR: pBACe3.6
           This sequence is the entire insert of clone RP11-544M22 The true
           left end of clone RP11-137A12 is at 156538 in this sequence. The
           true right end of clone RP11-381D2 is at 145015 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .200822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-544M22"
                     /clone_lib="RPCI-11.2"
     misc_feature    10395. .10397
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    44098. .44107
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    59585. .59604
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    90797. .90883
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    90861. .90883
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    55215 a  49052 c  46183 g  50372 t
ORIGIN

  Query Match             40.3%;  Score 571;  DB 9;  Length 200822;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  571;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174499 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 174558

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174559 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 174618

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174619 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 174678

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174679 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 174738

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174739 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 174798

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174799 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 174858

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174859 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 174918

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174919 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 174978

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174979 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 175038

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||
Db 175039 TGGCATTGAAAAATATAATAATCATAAAGTC 175069


RESULT 5
AX368386/c
LOCUS       AX368386                377 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1096 from Patent WO0204514.
ACCESSION   AX368386
VERSION     AX368386.1  GI:18856459
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1096 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       94 a    100 c    112 g     71 t
ORIGIN

  Query Match             23.0%;  Score 326;  DB 6;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-190;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy     189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy     249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy     309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy     369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy     429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18
```

```
Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1


RESULT 6
G13436/c
LOCUS       G13436                   441 bp    DNA     linear   STS 04-JUN-1996
DEFINITION  human STS WI-12338, sequence tagged site.
ACCESSION   G13436
VERSION     G13436.1  GI:1127545
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TATATTTTTCAATGCCAGTGCTG
            Primer B: AAACTTTCACCTGAGCTTCACC
            STS size: 130
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T78621).
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="750.8 cR from top of Chr1 linkage group"
     STS             16. .145
     primer_bind     16. .38
     primer_bind     complement(124. .145)
BASE COUNT      107 a     95 c    130 g    103 t      6 others
ORIGIN

  Query Match             12.2%;  Score 173;  DB 11;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 2e-95;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1178 CCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCT 1237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    216 CCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCT 157

Qy   1238 GCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 1297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 GCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 97

Qy   1298 TGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGG 1350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 TGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGG 44


RESULT 7
AX397469/c
LOCUS       AX397469                 2261 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1684 from Patent WO0212328.
ACCESSION   AX397469
VERSION     AX397469.1  GI:21068216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   King,G.E., Meagher,M.J., Xu,J. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0212328-A 1684 14-FEB-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2261
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      543 a    594 c    541 g    583 t
ORIGIN

  Query Match             6.1%;  Score 87;  DB 6;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-42;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 8
AK001497/c
LOCUS       AK001497                 2261 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ10635 fis, clone NT2RP2005669, highly similar
            to Homo sapiens mRNA for DEDD protein.
ACCESSION   AK001497
VERSION     AK001497.1  GI:7022790
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2005669.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
            Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,
            Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K.,
            Masuho,Y. and Kanehori,K.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2261)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
```

```
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2261
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2005669"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
BASE COUNT      543 a    594 c    541 g    583 t
ORIGIN

  Query Match             6.1%;  Score 87;  DB 9;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-42;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 9
G72919
LOCUS       G72919                   847 bp    DNA     linear   STS 08-AUG-2001
DEFINITION  MARC 2849-2850:991933517:1 SCF - porcine spleen Sus scrofa STS
            genomic, sequence tagged site.
ACCESSION   G72919
VERSION     G72919.1  GI:15146949
KEYWORDS    STS.
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 847)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
            Keele,J.W.
  TITLE     Single nucleotide polymorphism (SNP) discovery in expressed porcine
            genes
  JOURNAL   Unpublished
COMMENT
            Contact: Freking BA
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4278
            Fax: 402 762 4173
            Email: freking@email.marc.usda.gov
            Primer A: GGGTTCAGCTTTTGGATCTG
            Primer B: GCTGAGCCTGTGTTTCATCA
            STS size: 700
            PCR Profile:
                 Hotstart:  95 degrees for 15 minutes
                 Denature:  95 degrees for 30 seconds
                 Anneal:    58 degrees
                 Extension: 68 degrees for 2 minutes
                 Cycles:    32 to 45
            Protocol:
                 Template:       50-200 ng genomic DNA
                 Primer:         each 20 pmoles
                 dNTPs:          each 88 uM
                 Taq Polymerase: 0.25 units (Qiagen HotStar)

            Buffer:
                 Commercially supplied Qiagen HotStar buffer

            The STS is derived from PCR amplicons generated from genomic DNA,
            sequenced from each end using the amplification primers.  The
            sequence does not necessarily represent the entire amplicon.
            Sequence derived from PolyPhred was trimmed from each end of each
            unique contig until five consecutive bases exceeded a quality score
            threshold of 20, and the next 10 bases averaged a quality score of
            20 or greater. Amplicon size was estimated by agarose gel
            electrophoresis.
FEATURES             Location/Qualifiers
     source          1. .847
                     /organism="Sus scrofa"
                     /strain="white composite, duroc, meishan, minzhu,
                     fengjing, crossbreds"
                     /db_xref="taxon:9823"
                     /sex="male and female"
                     /clone_lib="SCF - porcine spleen"
                     /dev_stage="adult"
                     /note="Organ: spleen"
     STS             <1. .>847
BASE COUNT      205 a    249 c    200 g    193 t
ORIGIN

  Query Match             3.3%;  Score 47;  DB 11;  Length 847;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    873 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCA 919
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCA 578


RESULT 10
AF069988
LOCUS       AF069988                 1338 bp    mRNA    linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase 1 (Nit1) mRNA, complete cds.
ACCESSION   AF069988
VERSION     AF069988.1  GI:3228667
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1338
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            1. .1338
                     /gene="Nit1"
     CDS             58. .1029
```

```
                     /gene="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC40185.1"
                     /db_xref="GI:3228668"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      347 a    335 c    330 g    326 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 1338;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db    1269 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1301


RESULT 11
BC021634
LOCUS       BC021634               1365 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
            complete cds.
ACCESSION   BC021634
VERSION     BC021634.1  GI:18204912
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6754855.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:13825 IMAGE:4008543"
                     /tissue_type="Mammary tumor metastatized to lung.
                     MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
                     enhancer."
                     /clone_lib="NCI_CGAP_Lu30"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             59. .1030
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAH21634.1"
                     /db_xref="GI:18204913"
                     /db_xref="LocusID:27045"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYSSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      368 a    334 c    331 g    332 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 1365;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db    1270 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1302


RESULT 12
AF069985
LOCUS       AF069985               4481 bp    DNA     linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase homolog 1 (Nit1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069985
VERSION     AF069985.1  GI:3242979
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4481
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            541. .3944
                     /gene="Nit1"
     mRNA            join(541. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3944)
                     /gene="Nit1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            541. .606
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(605. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3658)
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
```

```
                     /protein_id="AAC40184.1"
                     /db_xref="GI:3242980"
                     /translation="MLGFITRPPHQLLCTGYRLLRIPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
     exon            863. .986
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=2
     exon            1177. .1263
                     /gene="Nit1"
                     /number=3
     exon            1638. .1889
                     /gene="Nit1"
                     /number=4
     exon            2015. .2118
                     /gene="Nit1"
                     /number=5
     exon            2362. .2495
                     /gene="Nit1"
                     /number=6
     exon            2626. .2751
                     /gene="Nit1"
                     /number=7
     exon            3392. .3944
                     /gene="Nit1"
                     /number=8
BASE COUNT     1101 a    1132 c    1133 g    1115 t
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 4481;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db   3898 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 3930


RESULT 13
AC105589
LOCUS       AC105589               181583 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
            ***, 67 unordered pieces.
ACCESSION   AC105589
VERSION     AC105589.2  GI:21736456
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 181583)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18092812.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNGI
                Center clone name: CH230-242B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 112684 bases at least Q40
                Consensus quality: 120012 bases at least Q30
                Consensus quality: 127587 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 67 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1312: contig of 1312 bp in length
            *     1313     1412: gap of unknown length
            *     1413     2762: contig of 1350 bp in length
            *     2763     2862: gap of unknown length
            *     2863     4412: contig of 1550 bp in length
            *     4413     4512: gap of unknown length
            *     4513     5532: contig of 1020 bp in length
            *     5533     5632: gap of unknown length
            *     5633     6814: contig of 1182 bp in length
```

```
*     6815     6914: gap of unknown length
*     6915     8099: contig of 1185 bp in length
*     8100     8199: gap of unknown length
*     8200     9786: contig of 1587 bp in length
*     9787     9886: gap of unknown length
*     9887    11574: contig of 1688 bp in length
*    11575    11674: gap of unknown length
*    11675    12816: contig of 1142 bp in length
*    12817    12916: gap of unknown length
*    12917    14067: contig of 1151 bp in length
*    14068    14167: gap of unknown length
*    14168    15370: contig of 1203 bp in length
*    15371    15470: gap of unknown length
*    15471    16926: contig of 1456 bp in length
*    16927    17026: gap of unknown length
*    17027    18088: contig of 1062 bp in length
*    18089    18188: gap of unknown length
*    18189    19315: contig of 1127 bp in length
*    19316    19415: gap of unknown length
*    19416    20772: contig of 1357 bp in length
*    20773    20872: gap of unknown length
*    20873    21937: contig of 1065 bp in length
*    21938    22037: gap of unknown length
*    22038    23147: contig of 1110 bp in length
*    23148    23247: gap of unknown length
*    23248    24803: contig of 1556 bp in length
*    24804    24903: gap of unknown length
*    24904    26111: contig of 1208 bp in length
*    26112    26211: gap of unknown length
*    26212    27647: contig of 1436 bp in length
*    27648    27747: gap of unknown length
*    27748    30257: contig of 2510 bp in length
*    30258    30357: gap of unknown length
*    30358    31725: contig of 1368 bp in length
*    31726    31825: gap of unknown length
*    31826    32912: contig of 1087 bp in length
*    32913    33012: gap of unknown length
*    33013    34593: contig of 1581 bp in length
*    34594    34693: gap of unknown length
*    34694    36338: contig of 1645 bp in length
*    36339    36438: gap of unknown length
*    36439    38139: contig of 1701 bp in length
*    38140    38239: gap of unknown length
*    38240    39349: contig of 1110 bp in length
*    39350    39449: gap of unknown length
*    39450    40968: contig of 1519 bp in length
*    40969    41068: gap of unknown length
*    41069    42068: contig of 1000 bp in length
*    42069    42168: gap of unknown length
*    42169    43869: contig of 1701 bp in length
*    43870    43969: gap of unknown length
*    43970    46141: contig of 2172 bp in length
*    46142    46241: gap of unknown length
*    46242    48158: contig of 1917 bp in length
*    48159    48258: gap of unknown length
*    48259    50613: contig of 2355 bp in length
*    50614    50713: gap of unknown length
*    50714    52725: contig of 2012 bp in length
*    52726    52825: gap of unknown length
*    52826    54534: contig of 1709 bp in length
*    54535    54634: gap of unknown length
*    54635    56455: contig of 1821 bp in length
*    56456    56555: gap of unknown length
*    56556    58494: contig of 1939 bp in length
*    58495    58594: gap of unknown length
*    58595    60496: contig of 1902 bp in length
*    60497    60596: gap of unknown length
*    60597    62628: contig of 2032 bp in length
*    62629    62728: gap of unknown length
*    62729    65060: contig of 2332 bp in length
*    65061    65160: gap of unknown length
*    65161    66543: contig of 1383 bp in length
*    66544    66643: gap of unknown length
*    66644    68412: contig of 1769 bp in length
*    68413    68512: gap of unknown length
*    68513    71585: contig of 3073 bp in length
*    71586    71685: gap of unknown length
*    71686    74575: contig of 2890 bp in length
*    74576    74675: gap of unknown length
*    74676    77865: contig of 3190 bp in length
*    77866    77965: gap of unknown length
*    77966    80442: contig of 2477 bp in length
*    80443    80542: gap of unknown length
*    80543    83108: contig of 2566 bp in length
*    83109    83208: gap of unknown length
*    83209    86895: contig of 3687 bp in length
*    86896    86995: gap of unknown length
*    86996    91038: contig of 4043 bp in length
*    91039    91138: gap of unknown length
*    91139    94412: contig of 3274 bp in length
*    94413    94512: gap of unknown length
*    94513    96959: contig of 2447 bp in length
*    96960    97059: gap of unknown length
*    97060    98988: contig of 1929 bp in length
*    98989    99088: gap of unknown length
*    99089   101633: contig of 2545 bp in length
*   101634   101733: gap of unknown length
*   101734   105899: contig of 4166 bp in length

  Query Match             2.3%;  Score 33;  DB 2;  Length 181583;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db 135691 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 135723


RESULT 14
AC084821
LOCUS       AC084821              211772 bp    DNA     linear   ROD 19-FEB-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-395h6, complete sequence.
ACCESSION   AC084821
VERSION     AC084821.25  GI:18702388
KEYWORDS    HTG.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-395h6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 19, 2002 this sequence version replaced gi:18390260.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .211772
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
```

```
                     /clone="rp23-395h6"
                     /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT    55607 a  48743 c  49203 g  58215 t      4 others
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 211772;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db   3237 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 3269


RESULT 15
AC087229/c
LOCUS       AC087229              215043 bp    DNA     linear   ROD 14-AUG-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,
            complete sequence.
ACCESSION   AC087229
VERSION     AC087229.24  GI:22094398
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Mus musculus Chromosome 1 BAC Clone rp23-191a19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Aug 3, 2002 this sequence version replaced gi:22038565.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .215043
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-191a19"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT    55806 a  49340 c  51162 g  58732 t      3 others
ORIGIN

  Query Match             2.3%;  Score 33;  DB 10;  Length 215043;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1343 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 1375
           |||||||||||||||||||||||||||||||||
Db  87663 TGCTGAGGGCTGAGCAGCACTGGCATTGAAAAA 87631


Search completed: November 30, 2002, 13:41:10
Job time : 3446 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:36:49 ; Search time 253 Seconds
(without alignments)
12604.070 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1314	92.8	1416	21	AAZ46101	Nucleotide sequenc
	2	1160	81.9	1382	21	AAF16257	Human prostate can
	3	766	54.1	1203	20	AAX30398	DNA encoding a hum
c	4	326	23.0	377	24	ABK39058	cDNA encoding lung
	5	91	6.4	505	22	AAH07371	Human cDNA clone (
	6	87	6.1	592	22	AAH11324	Human cDNA clone (
c	7	87	6.1	2261	22	AAH14367	Human cDNA sequenc
c	8	87	6.1	2261	24	ABK46133	cDNA encoding colo
	9	58	4.1	60	24	ABN39628	Human spliced tran
	10	52	3.7	539	22	AAK61236	Human immune/haema
	11	30	2.1	3300	20	AAZ11915	Human potassium ch
	12	29	2.0	80	24	ABA82722	Human protective D
	13	29	2.0	223	23	ABV08728	Human prostate exp
c	14	29	2.0	311	24	ABL82990	Human ovarian canc
c	15	29	2.0	384	23	ABV13242	Human prostate exp
c	16	29	2.0	406	23	ABV04073	Human prostate exp
c	17	29	2.0	416	23	ABV34361	Human prostate exp
	18	29	2.0	417	23	ABV58938	Human prostate exp
	19	29	2.0	528	23	ABV38626	Human prostate exp
	20	29	2.0	563	20	AAX98006	Human secreted pro
c	21	29	2.0	564	24	ABN93758	Gene #256 used to
c	22	29	2.0	748	22	AAK63693	Human immune/haema
	23	29	2.0	807	24	ABA82714	Human protective D
	24	29	2.0	1193	13	AAQ30670	GGF2BPP2.CDS. Syn
	25	29	2.0	1193	15	AAQ58303	GGF-II cDNA sequen
	26	29	2.0	1193	15	AAQ62849	GGF-II cDNA sequen
	27	29	2.0	1193	16	AAQ74885	Putative bovine gl
	28	29	2.0	1193	17	AAT48079	Bovine neuregulin
	29	29	2.0	1193	17	AAT30997	Bovine glial growt
	30	29	2.0	1193	17	AAT06703	Bovine glial growt
	31	29	2.0	1211	22	AAH31352	Human secreted pro
	32	29	2.0	1608	21	AAC95501	Human secreted pro
	33	29	2.0	2317	22	AAS03917	Human secreted pro
	34	29	2.0	3625	23	AAF87120	NOV9 coding sequen
	35	29	2.0	4843	20	AAX37553	Human HLIG-1 cDNA.
c	36	28	2.0	319	24	ABL79372	Human ovarian canc
	37	28	2.0	324	23	ABV49621	Human prostate exp
	38	28	2.0	326	23	ABV19858	Human prostate exp
	39	28	2.0	428	23	ABV17820	Human prostate exp
	40	28	2.0	469	23	ABV47612	Human prostate exp
c	41	28	2.0	619	22	AAL22931	Human breast cance
c	42	28	2.0	711	22	AAI97858	Human neuroblastom
	43	28	2.0	941	24	ABN98591	Arabidopsis thalia
	44	28	2.0	1108	13	AAQ30671	GGF2BPP3.CDS. Syn
	45	28	2.0	1108	15	AAQ58304	GGF-II cDNA sequen

ALIGNMENTS

RESULT 1
AAZ46101
ID AAZ46101 standard; DNA; 1416 BP.
XX
AC AAZ46101;
XX
DT 05-MAY-2000 (first entry)
XX
DE Nucleotide sequence of the coding region of NIT1 gene.
XX
KW NIT1 gene; nitrilase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW FRA3B; cancer; genome allele inactivation; ss.
XX
OS Homo sapiens.
OS Mus sp.
OS Drosophila melanogaster.
OS Caenorhabditis elegans.
XX
FH Key Location/Qualifiers
FT CDS 3..1415
FT /*tag= a
FT /product= (pos: 18..20, aa: Xaa)
FT /product= (pos: 1179..1181, aa: Gln)
FT /product= (pos: 1182..1184, aa: Ala)
FT /note= "contains 7 internal stop codons; Xaa is an
FT unspecified amino acid"
XX
PN WO200003685-A2.
XX
PD 27-JAN-2000.
XX
PF 20-JUL-1999; 99WO-US16366.

```
XX
PR   20-JUL-1998;   98US-0093350.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce CM;
XX
DR   WPI; 2000-171195/15.
DR   P-PSDB; AAY68739.
XX
PT   Novel nitrilase homologs used as diagnostic and therapeutic reagents
PT   for the detection and treatment of cancer -
XX
PS   Claim 6; Fig 6; 25pp; English.
XX
CC   The present sequence represents the coding region of human, murine,
CC   Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC   human and mouse NIT1 genes are members of an uncharacterised
CC   mammalian gene family with homology to bacterial and plant nitrilases.
CC   The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC   for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC   In mouse and humans, FHIT and NIT are encoded by two different genes,
CC   localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC   human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC   chromosomal fragile site FRA3B, is often altered in most common forms
CC   of human cancer. The Nit1 protein overcomes the mutated inactivation
CC   of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC   and analogues of them, and antibodies are used as diagnostic and
CC   therapeutic reagents for the detection and treatment of cancers.
XX
SQ   Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

  Query Match             92.8%;  Score 1314;  DB 21;  Length 1416;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1414;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCACTCGCTGCGGCCTNTCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCCCACTCGCTGCGGCCTNTCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGG 60

Qy      61 TTTTGGCTATATCTTCATGTAGGACCTACTCCCTATCCCGTCGGCCGCGGCTGGGCTTCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTTGGCTATATCTTCATGTAGGACCTACTCCCTATCCCGTCGGCCGCGGCTGGGCTTCA 120

Qy     121 TCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAAC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAAC 180

Qy     181 TCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCG 240

Qy     241 AACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTA 300

Qy     301 AAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360

Qy     361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCAC 420

Qy     421 TGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGT 480

Qy     481 CCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATT 540

Qy     541 GTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGT 600

Qy     601 GTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660

Qy     661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATG 720

Qy     721 ACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCT 780

Qy     781 ATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCC 840

Qy     841 GTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGA 900

Qy     901 AGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

Qy     961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGC 1020

Qy    1021 GCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACC 1080

Qy    1081 CACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCC 1140

Qy    1141 ACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTG 1200
           ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db    1141 ACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCACCCACAGCTCCCCTCACTTG 1200

Qy    1201 GAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCT 1260

Qy    1261 GAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTT 1320

Qy    1321 TATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAA 1330

Qy    1381 TAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||||||||
Db    1381 TAATCATAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416


RESULT 2
AAF16257
ID   AAF16257 standard; cDNA; 1382 BP.
XX
AC   AAF16257;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
```

```
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB57054.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 1124; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

  Query Match             81.9%;  Score 1160;  DB 21;  Length 1382;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1280;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy     170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy     230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy     290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318

Qy     350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy     410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy     470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy     530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy     590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGG-CTATGTGTGAAAGCAACTCTAC 617

Qy     650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     618 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 677

Qy     710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 737

Qy     770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 797

Qy     830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 857

Qy     890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 917

Qy     950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 977

Qy    1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1037

Qy    1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1097

Qy    1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1157

Qy    1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1217

Qy    1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1277

Qy    1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1337

Qy    1370 GAAAAATATAATAATCATAAA 1390
           ||||||||||||||||||||
Db    1338 GAAAAATATAATAATCATAAA 1358


RESULT 3
AAX30398
ID   AAX30398 standard; DNA; 1203 BP.
XX
AC   AAX30398;
XX
DT   14-MAY-1999  (first entry)
XX
```

DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN WO9907891-A1.
XX
PD 18-FEB-1999.
XX
PF 04-AUG-1998; 98WO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX
DR WPI; 1999-167452/14.
DR P-PSDB; AAY10877.
XX
PT New isolated human genes encoding secreted polypeptides - useful for
PT diagnosis and treatment of pathalogical diseases
XX
PS Claim 3; Page 265-266; 331pp; English.
XX
CC The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the amount of the
CC secreted polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
XX
SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

Query Match 54.1%; Score 766; DB 20; Length 1203;
Best Local Similarity 100.0%; Pred. No. 8.4e-305;
Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    628 CTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACAC 687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 CTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACAC 220

Qy    688 CAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGG 747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 CAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGG 280

Qy    748 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 340

Qy    808 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAG 867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAG 400

Qy    868 TGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGG 927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 TGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGG 460

Qy    928 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTG 987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTG 520

Qy    988 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACC 1047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACC 580

Qy   1048 GCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGA 1107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 GCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGA 640

Qy   1108 GTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTG 1167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 GTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTG 700

Qy   1168 GAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 1227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    701 GAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 760

Qy   1228 CAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGT 1287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    761 CAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGT 820

Qy   1288 TTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTG 1347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    821 TTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTG 880

Qy   1348 AGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||||||||||||||||||||
Db    881 AGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 926
```

RESULT 4
ABK39058/c
ID ABK39058 standard; cDNA; 377 BP.
XX
AC ABK39058;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA encoding lung tumour protein clone R0130:B11.
XX
KW Lung tumour; cancer; T cell; immune response stimulator;
KW cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200204514-A2.

```
XX
PD   17-JAN-2002.
XX
PF   10-JUL-2001; 2001WO-US22058.
XX
PR   11-JUL-2000; 2000US-0614124.
PR   29-AUG-2000; 2000US-0651563.
PR   08-SEP-2000; 2000US-0658824.
PR   26-SEP-2000; 2000US-0671325.
PR   06-OCT-2000; 2000US-0677419.
PR   30-OCT-2000; 2000US-0702705.
PR   13-DEC-2000; 2000US-0736457.
PR   03-MAY-2001; 2001US-0849626.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Watanabe Y,  Henderson RA,  Johnson JC,  Retter MW;
PI   Marnerakis M,  Carter D,  Fanger GR,  Vedvick TS,  Bangur CS;
PI   Mcnabb A,  Wang A,  Fanger N,  Switzer A,  Mcneill PD,  Clapper JD;
XX
DR   WPI; 2002-164634/21.
XX
PT   Novel polynucleotide encoding a lung tumour polypeptide useful for
PT   stimulating and/or expanding T cells specific for a tumour protein -
XX
PS   Claim 1; SEQ ID No 1096; 223pp; English.
XX
CC   The invention describes an isolated polynucleotide and polypeptide
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein for determining the presence of a cancer in a patient. A
CC   composition containing the polynucleotide and/or polypeptide is useful
CC   for treating a lung cancer in a patient. The polypeptide is useful for
CC   removing tumour cells from a biological sample. The polynucleotide is
CC   also useful as probe or primer to detect the level of mRNA encoding a
CC   tumour protein. This sequence encodes a lung tumour associated protein
CC   or protein fragment, described in the method of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 377 BP; 94 A; 100 C; 112 G; 71 T; 0 other;

  Query Match             23.0%;  Score 326;  DB 24;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 3.7e-124;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy    189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy    249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy    309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy    369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy    429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1


RESULT 5
AAH07371
ID   AAH07371 standard; cDNA; 505 BP.
XX
AC   AAH07371;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:4206.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 4206; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 505 BP; 105 A; 120 C; 167 G; 108 T; 5 other;

  Query Match              6.4%;  Score 91;  DB 22;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 1e-27;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     20 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCATG 79
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCATG 78

Qy      80 TAGGACCTACTCCCTATCCCGTCGGCCGCGG 110
           ||||||||||||||||||||||||||||||
Db      79 TAGGACCTACTCCCTATCCCGTCGGCCGCGG 109


RESULT 6
AAH11324
ID   AAH11324 standard; cDNA; 592 BP.
XX
AC   AAH11324;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:8159.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 8159; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 592 BP; 164 A; 134 C; 138 G; 148 T; 8 other;

  Query Match             6.1%;  Score 87;  DB 22;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 60

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           ||||||||||||||||||||||||||
Db     61 ATTGAAAAATATAATAATCATAAAGTC 87


RESULT 7
AAH14367/c
ID   AAH14367 standard; cDNA; 2261 BP.
XX
AC   AAH14367;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11772.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11772; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
```

```
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

  Query Match             6.1%;  Score 87;  DB 22;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 8
ABK46133/c
ID   ABK46133 standard; cDNA; 2261 BP.
XX
AC   ABK46133;
XX
DT   05-JUN-2002  (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 1684.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic;
KW   immunotherapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US24218.
XX
PR   03-AUG-2000; 2000US-223283P.
PR   28-MAR-2001; 2001US-279763P.
PR   29-JUN-2001; 2001US-302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE,  Meagher MJ,  Xu J,  Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers
PT   for the progression of cancer -
XX
PS   Claim 1; SEQ ID No 1684; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer.
CC   ABK44450-ABK46237 represent coding sequences of human colon tumour
CC   proteins of the invention.
CC   Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC   for this patent did not form part of the printed specification but was
CC   supplied by the European Patent Office.
XX
SQ   Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

  Query Match             6.1%;  Score 87;  DB 24;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 9
ABN39628
ID   ABN39628 standard; DNA; 60 BP.
XX
AC   ABN39628;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human spliced transcript detection oligonucleotide SEQ ID NO:12376.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB01903.
XX
PR   28-JUL-2000; 2000US-221607P.
PR   02-MAY-2001; 2001US-287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A,  Wasserman A,  Mintz E,  Mintz L,  Faigler S;
XX
DR   WPI; 2002-257383/30.
XX
PT   New oligonucleotide libraries comprising oligonucleotides which
PT   selectively hybridize to mRNAs transcribed from a transcription unit of
PT   a genome, useful for detecting tissue-, pathology-, and
PT   developmental-specific genes -
XX
PS   Example 1; SEQ ID 12376; 47pp; English.
XX
CC   The present invention describes oligonucleotide libraries for detecting
CC   messenger RNAs that populate a (sub-)transcriptome, where the
CC   (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC   transcription units that populate a genome. The library comprises
CC   several oligonucleotides, each capable of hybridising selectively to a
CC   set of messenger RNAs transcribed from a given transcription unit of
CC   the genome, which encodes one or more messenger RNA splice variants.
CC   The oligonucleotide libraries are useful for detecting mRNAs from a
CC   biological sample, in expression profiling studies, in qualitatively or
CC   quantitatively characterising the corresponding transcriptome, and in
CC   detecting RNA transcripts and splice variants of human or animal
CC   transcriptomes. The libraries may also be used as specialised mini
CC   libraries to detect transcripts of a sub-transcriptome under a
CC   particular biological or pathological state, and so allowing the
CC   detection of tissue- and pathology-specific genes such as those genes
CC   only expressed in specific tissue under a specific pathological
CC   condition; to detect developmental specific genes; and to detect RNA
CC   transcripts and splice variants of a transcriptome of a patient suffering
CC   from a particular disorder. ABN27253 to ABN59589 represent
CC   oligonucleotide sequences from rats, humans and mice, which are used in
CC   the exemplification of the present invention.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence 60 BP; 9 A; 20 C; 14 G; 17 T; 0 other;

  Query Match             4.1%;  Score 58;  DB 24;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   58;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCA 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 TCTGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATATCTTCA 60


RESULT 10
AAK61236
ID   AAK61236 standard; cDNA; 539 BP.
XX
AC   AAK61236;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6296.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
```

```
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM88455.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 1; SEQ ID NO 6296; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

  Query Match             3.7%;  Score 52;  DB 22;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 1e-11;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     109 GGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 GGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 422


RESULT 11
AAZ11915
ID   AAZ11915 standard; cDNA; 3300 BP.
XX
AC   AAZ11915;
XX
DT   30-NOV-1999  (first entry)
XX
DE   Human potassium channel K+Hnov59 cDNA.
XX
KW   Potassium channel; ataxia; arrhythmia; epilepsy; Bartter's syndrome;
KW   cardiovascular disorder; CNS disorder; renal disorder; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             50..1285
FT                   /*tag=  a
FT                   /product= "Human K+Hnov59 potassium channel"
XX
PN   WO9943696-A1.
XX
PD   02-SEP-1999.
XX
PF   22-FEB-1999;   99WO-US03826.
XX
PR   19-JAN-1999;   99US-0116448.
PR   25-FEB-1998;   98US-0076687.
PR   07-AUG-1998;   98US-0095836.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Curran ME,  Hu P,  Miller AP,  Rutter M,  Wang J;
XX
DR   WPI; 1999-527591/44.
DR   P-PSDB; AAY34133.
XX
PT   New nucleic acids encoding mammalian K+Hnov potassium channel
PT   proteins, useful for the diagnosis and treatment of episodic ataxia
PT   with myokymia, cardiac arrhythmia, epilepsy and Bartter's syndrome
XX
PS   Claim 4; Page 102-104; 112pp; English.
XX
CC   This sequence represents human potassium channel K+Hnov59 cDNA.
CC   K+Hnov proteins have a high degree of homology to known potassium
CC   channels and may be alpha subunits, which form the functional channel,
CC   or accessory subunits that act to modulate the channel activity. K+Hnov59
CC   is a 4 transmembrane domain, 2 pore domain potassium channel. The gene
CC   is located on chromosome 19, determined via PCR chromosomal
CC   localisation using primers AAZ11939 and AAZ11940. K+Hnov cDNAs
CC   were isolated by extension of expressed sequence tags (ESTs) which were
CC   related but not identical to known human potassium channels. Potential
CC   polymorphisms detected as sequence variants between multiple
CC   independent clones. Potassium channels have critical roles in various
CC   cell types and biochemical pathways. Defective potassium channels are
CC   known to cause four human diseases: episodic ataxia with myokymia;
CC   cardiac arrhythmia (long QT syndrome); epilepsy; and Bartter's syndrome.
CC   As potassium channels are critical components of virtually all cells,
CC   it is likely that abnormal potassium channels are also implicated in
CC   certain renal, cardiovascular and central nervous system (CNS)
CC   disorders. Nucleotides encoding K+Hnov proteins may be used for
CC   identifying homologous or related proteins and the DNA sequences encoding
CC   them. They may be used to produce compositions that modulate the
CC   expression and function of the K+Hnov protein and in studying the
CC   biochemical pathways associated with it. They may also be used for the
CC   recombinant production of K+Hnov protein in fermentation cultures.
CC   Additionally, such nucleotides may be used in gene therapy protocols for
CC   the treatment of diseases associated with abnormal potassium channels.
XX
SQ   Sequence 3300 BP; 997 A; 629 C; 680 G; 994 T; 0 other;

  Query Match             2.1%;  Score 30;  DB 20;  Length 3300;
  Best Local Similarity   100.0%;  Pred. No. 0.0093;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||||
Db    3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAAA 3267
```

```
RESULT 12
ABA82722
ID   ABA82722 standard; DNA; 80 BP.
XX
AC   ABA82722;
XX
DT   07-FEB-2002  (first entry)
XX
DE   Human protective DNA sequence CNI-00735 fragment #8.
XX
KW   Human; protective sequence; cell death; cancer; autoimmune disease;
KW   neurological disorder; stroke; cytostatic; neuroprotective; gene therapy;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200176457-A2.
XX
PD   18-OCT-2001.
XX
PF   09-APR-2001; 2001WO-US11663.
XX
PR   11-APR-2000; 2000US-0547735.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Thomas MB,  Portbury SD,  Puranam K,  Katz LC,  Lo DC,  Barney S;
XX
DR   WPI; 2002-025874/03.
DR   P-PSDB; ABB44640.
XX
PT   New protective sequences and their products, useful for diagnosing and
PT   treating diseases involving cell death, including neurological
PT   disorders e.g. stroke and for identifying modulators of expression of
PT   the protective sequences  -
XX
PS   Claim 2; Fig 5; 283pp; English.
XX
CC   The present invention relates to protective sequence proteins
CC   (ABB44624-ABB44830) and their coding sequences (ABA82701-ABA82937).
CC   The sequences, when introduced into a cell either predisposed to undergo
CC   cell death or in the process of undergoing cell death, prevent, delay or
CC   rescue the cell from death, hence, these sequences are named "protective
CC   sequences". The sequences are useful for treating and/or ameliorating
CC   cancer, autoimmune diseases and neurological disorders e.g. stroke.
CC   Further examples of diseases which may be treated by the present
CC   invention are given in the specification.
XX
SQ   Sequence 80 BP; 40 A; 10 C; 15 G; 15 T; 0 other;

  Query Match             2.0%;  Score 29;  DB 24;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 0.035;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db     39 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 67


RESULT 13
ABV08728
ID   ABV08728 standard; cDNA; 223 BP.
XX
AC   ABV08728;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 8719.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 1376; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 223 BP; 76 A; 44 C; 64 G; 38 T; 1 other;

  Query Match             2.0%;  Score 29;  DB 23;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 0.031;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db    188 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 216


RESULT 14
ABL82990/c
ID   ABL82990 standard; cDNA; 311 BP.
XX
AC   ABL82990;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:5968.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US17756.
XX
```

PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
PS Claim 1; SEQ ID 5968; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 311 BP; 96 A; 42 C; 45 G; 128 T; 0 other;

Query Match 2.0%; Score 29; DB 24; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 40 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 12

RESULT 15
ABV13242/c
ID ABV13242 standard; cDNA; 384 BP.
XX
AC ABV13242;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 13233.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 2192; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 384 BP; 136 A; 69 C; 73 G; 106 T; 0 other;

Query Match 2.0%; Score 29; DB 23; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 42 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 14

Search completed: November 30, 2002, 12:43:36
Job time : 267 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:38:59 ; Search time 58 Seconds
(without alignments)
7487.143 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	2.1	3300	4	US-09-336-643A-82	Sequence 82, Appl
2	29	2.0	1193	1	US-08-036-555B-134	Sequence 134, App
3	29	2.0	1193	1	US-08-469-569-134	Sequence 134, App
4	29	2.0	1193	1	US-08-249-322A-134	Sequence 134, App
5	29	2.0	1193	1	US-08-469-526A-134	Sequence 134, App
6	29	2.0	1193	2	US-08-734-591A-134	Sequence 134, App
7	29	2.0	1193	2	US-08-469-660-134	Sequence 134, App
8	29	2.0	1193	3	US-08-341-018-3	Sequence 3, Appli
9	29	2.0	1193	3	US-08-470-335-134	Sequence 134, App
10	29	2.0	1193	4	US-08-735-021-134	Sequence 134, App
11	29	2.0	1193	4	US-08-734-664A-134	Sequence 134, App
12	29	2.0	1193	4	US-08-470-339-134	Sequence 134, App
13	29	2.0	1193	4	US-08-467-602-134	Sequence 134, App
14	29	2.0	1193	5	PCT-US94-05083C-130	Sequence 130, App
15	29	2.0	1193	5	PCT-US95-06846A-134	Sequence 134, App
16	29	2.0	4843	3	US-08-986-485-1	Sequence 1, Appli
17	28	2.0	1108	1	US-08-036-555B-135	Sequence 135, App
18	28	2.0	1108	1	US-08-469-569-135	Sequence 135, App
19	28	2.0	1108	1	US-08-249-322A-135	Sequence 135, App
20	28	2.0	1108	1	US-08-469-526A-135	Sequence 135, App
21	28	2.0	1108	2	US-08-734-591A-135	Sequence 135, App
22	28	2.0	1108	2	US-08-469-660-135	Sequence 135, App
23	28	2.0	1108	3	US-08-341-018-5	Sequence 5, Appli
24	28	2.0	1108	3	US-08-470-335-135	Sequence 135, App
25	28	2.0	1108	4	US-08-735-021-135	Sequence 135, App
26	28	2.0	1108	4	US-08-734-664A-135	Sequence 135, App
27	28	2.0	1108	4	US-08-470-339-135	Sequence 135, App
28	28	2.0	1108	4	US-08-467-602-135	Sequence 135, App
29	28	2.0	1108	5	PCT-US94-05083C-131	Sequence 131, App
30	28	2.0	1108	5	PCT-US95-06846A-135	Sequence 135, App
31	28	2.0	3238	4	US-08-123-934A-5	Sequence 5, Appli
32	28	2.0	3238	5	PCT-US94-10080-5	Sequence 5, Appli
33	28	2.0	8201	1	US-08-253-155A-9	Sequence 9, Appli
34	27	1.9	989	2	US-08-874-460-1	Sequence 1, Appli
35	27	1.9	1969	2	US-08-541-033A-7	Sequence 7, Appli
36	27	1.9	1969	2	US-08-828-451-7	Sequence 7, Appli
37	27	1.9	2096	2	US-08-541-033A-19	Sequence 19, Appl
38	27	1.9	2096	2	US-08-828-451-19	Sequence 19, Appl
39	27	1.9	2099	2	US-08-541-033A-3	Sequence 3, Appli
40	27	1.9	2099	2	US-08-828-451-3	Sequence 3, Appli
41	27	1.9	2137	2	US-08-541-033A-18	Sequence 18, Appl
42	27	1.9	2137	2	US-08-828-451-18	Sequence 18, Appl
43	27	1.9	2140	2	US-08-541-033A-1	Sequence 1, Appli
44	27	1.9	2140	2	US-08-828-451-1	Sequence 1, Appli
45	27	1.9	2420	5	PCT-US93-00227-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-336-643A-82
; Sequence 82, Application US/09336643A
; Patent No. 6399761
; GENERAL INFORMATION:
; APPLICANT: Miller, Andrew P.
; APPLICANT: Curran, Mark Edward
; APPLICANT: Hu, Ping
; APPLICANT: Rutter, Marc
; APPLICANT: Wang, Jian-Wang
; TITLE OF INVENTION: No. 6399761el Human Potassium Channels
; FILE REFERENCE: SEQ-15P
; CURRENT APPLICATION NUMBER: US/09/336,643A
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 60/076,687
; PRIOR FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: 60/116,448
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: PCT/US99/03826
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 3300
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1285)
US-09-336-643A-82

Query Match 2.1%; Score 30; DB 4; Length 3300;
Best Local Similarity 100.0%; Pred. No. 0.00092;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
||||||||||||||||||||||||||||||
Db 3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 3267

RESULT 2
US-08-036-555B-134
; Sequence 134, Application US/08036555B
; Patent No. 5530109
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use

```
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/036,555B
;      FILING DATE:  24-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APRIL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.K. 91 07566.3
;      FILING DATE:  10-APRIL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsai, Christine H.
;      REGISTRATION NUMBER:  34,266
;      REFERENCE/DOCKET NUMBER:  LUD 5250.4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-036-555B-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 3
US-08-469-569-134
; Sequence 134, Application US/08469569
; Patent No. 5606032
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew; Stroobant, Paul;
;    APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;    APPLICANT:  Chen, Maio Su; Hiles, Ian
;    TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;    TITLE OF INVENTION:  Preparation and Use
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,569
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  24-MAR-1993
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APRIL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.K. 91 07566.3
;      FILING DATE:  10-APRIL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsai, Christine H.
;      REGISTRATION NUMBER:  34,266
;      REFERENCE/DOCKET NUMBER:  LUD 5250.4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-469-569-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 4
US-08-249-322A-134
; Sequence 134, Application US/08249322A
; Patent No. 5716930
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew; Stroobant, Paul;
;    APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;    APPLICANT:  Chen, Maio Su; Hiles, Ian
;    TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;    TITLE OF INVENTION:  Preparation and Use
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
```

```
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/249,322A
;       FILING DATE:  26-MAY-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/036,555
;       FILING DATE:  24-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/965,173
;       FILING DATE:  23-OCT-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/940,389
;       FILING DATE:  03-SEP-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/907,138
;       FILING DATE:  30-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/863,703
;       FILING DATE:  03-APRIL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.K. 91 07566.3
;       FILING DATE:  10-APRIL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Tsai, Christine H.
;       REGISTRATION NUMBER:  34,266
;       REFERENCE/DOCKET NUMBER:  LUD 250.4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  134:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1193
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-249-322A-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 5
US-08-469-526A-134
; Sequence 134, Application US/08469526A
; Patent No. 5792849
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Maio Su
;    APPLICANT:  Hiles, Ian
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,526A
;       FILING DATE:  06 June 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/036,555
;       FILING DATE:  24-MAR-1993
;       APPLICATION NUMBER:  07/965,173
;       FILING DATE:  23-OCT-1992
;       APPLICATION NUMBER:  07/940,389
;       FILING DATE:  03-SEP-1992
;       APPLICATION NUMBER:  07/907,138
;       FILING DATE:  03-JUN-1992
;       APPLICATION NUMBER:  07/863,703
;       FILING DATE:  03-APRIL-1992
;       APPLICATION NUMBER:  U.K. 91 07566.3
;       FILING DATE:  10-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bieker-Brady, Kristina
;       REGISTRATION NUMBER:  39,109
;       REFERENCE/DOCKET NUMBER:  04585/00200A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-428-0200
;       TELEFAX:  617-428-7045
;   INFORMATION FOR SEQ ID NO:  134:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1193
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-469-526A-134

  Query Match             2.0%;  Score 29;  DB 1;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 6
US-08-734-591A-134
; Sequence 134, Application US/08734591A
; Patent No. 5854220
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Hiles, Ian
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Mario
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM Compatible Pentium
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  WordPerfect (Version 7.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,591A
;      FILING DATE:  22-OCT-1996
;      CLASSIFICATION:  536
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/470,335
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  03-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  UK 91 07566.3
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  04585/00200P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-734-591A-134

  Query Match             2.0%;  Score 29;  DB 2;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 7
US-08-469-660-134
; Sequence 134, Application US/08469660
; Patent No. 5876973
;  GENERAL INFORMATION:
;    APPLICANT:  Gwynne, David I.; Marchionni, Mark;
;    APPLICANT:  McBurney, Robert N.
;    TITLE OF INVENTION:  INHIBITORS OF CELL PROLIFERATION,
;    TITLE OF INVENTION:  THEIR PREPARATION AND USE
;    NUMBER OF SEQUENCES:  184
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      ZIP:  0211-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,660
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/011,396
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/984,085
;      FILING DATE:  01-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/951,747
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/927,337
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/017004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  200154
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-469-660-134

  Query Match             2.0%;  Score 29;  DB 2;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 8
US-08-341-018-3
; Sequence 3, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
;  APPLICANT: Gywnne, David I.
;  APPLICANT: Mahanthappa, Nagesh K.
;  APPLICANT: Marchionni, Mark A.
;  APPLICANT: Bermingham-McDonogh, Olivia
;  APPLICANT: Goldin, Stanley M.
;  APPLICANT: McBurney, Robert N.
;  TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
;  TITLE OF INVENTION: CELLULAR COMMUNICATION
;  FILE REFERENCE: 04585/041001
;  CURRENT APPLICATION NUMBER: US/08/341,018A
;  CURRENT FILING DATE: 1994-11-17
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-341-018-3

  Query Match             2.0%;  Score 29;  DB 3;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 9
US-08-470-335-134
; Sequence 134, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
```

```
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/00200B
;  CURRENT APPLICATION NUMBER: US/08/470,335F
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  NUMBER OF SEQ ID NOS: 252
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-470-335-134

  Query Match             2.0%;  Score 29;  DB 3;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 10
US-08-735-021-134
; Sequence 134, Application US/08735021B
; Patent No. 6194377
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/00200L
;  CURRENT APPLICATION NUMBER: US/08/735,021B
;  CURRENT FILING DATE: 1996-10-22
;  EARLIER APPLICATION NUMBER: 08/472,065
;  EARLIER FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  EARLIER APPLICATION NUMBER: 07/965,173
;  EARLIER FILING DATE: 1992-10-23
;  EARLIER APPLICATION NUMBER: 07/940,389
;  EARLIER FILING DATE: 1992-09-03
;  EARLIER APPLICATION NUMBER: 07/907,138
;  EARLIER FILING DATE: 1992-06-30
;  EARLIER APPLICATION NUMBER: 07/863,703
;  EARLIER FILING DATE: 1992-04-03
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (8)...(796)
US-08-735-021-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 11
US-08-734-664A-134
; Sequence 134, Application US/08734664A
; Patent No. 6204241
;  GENERAL INFORMATION:
;    APPLICANT:  Goodearl, Andrew
;    APPLICANT:  Stroobant, Paul
;    APPLICANT:  Minghetti, Luisa
;    APPLICANT:  Waterfield, Michael
;    APPLICANT:  Marchionni, Mark
;    APPLICANT:  Chen, Mario
;    APPLICANT:  Hiles, Ian
;    TITLE OF INVENTION:  GLIAL MITOGENIC FACTORS, THEIR
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  187
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM Compatible Pentium
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSeq Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,664A
;      FILING DATE:  22-OCT-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/249,322
;      FILING DATE:  26-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/036,555
;      FILING DATE:  24-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,173
;      FILING DATE:  23-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/940,389
;      FILING DATE:  03-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,138
;      FILING DATE:  30-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/863,703
;      FILING DATE:  03-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  UK 91 07566.3
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  04585/00200J
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  134:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-734-664A-134
```

```
  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 12
US-08-470-339-134
; Sequence 134, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
;  APPLICANT: GOODEARL, ANDREW
;  APPLICANT: STROOBANT, PAUL
;  APPLICANT: MINGHETTI, LUISA
;  APPLICANT: WATERFIELD, MICHAEL
;  APPLICANT: MARCHIONNI, MARK
;  APPLICANT: CHEN, MARIO S.
;  APPLICANT: HILES, IAN
;  TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
;  TITLE OF INVENTION: PREPARATION AND USE
;  FILE REFERENCE: 04585/002008
;  CURRENT APPLICATION NUMBER: US/08/470,339C
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/036,555
;  EARLIER FILING DATE: 1993-03-24
;  EARLIER APPLICATION NUMBER: 07/940,389
;  EARLIER FILING DATE: 1992-09-03
;  EARLIER APPLICATION NUMBER: 07/907,138
;  EARLIER FILING DATE: 1992-06-30
;  EARLIER APPLICATION NUMBER: 07/863,703
;  EARLIER FILING DATE: 1992-04-03
;  EARLIER APPLICATION NUMBER: 91 07566.3 GB
;  EARLIER FILING DATE: 1999-04-10
;  NUMBER OF SEQ ID NOS: 226
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-470-339-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 13
US-08-467-602-134
; Sequence 134, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
;  APPLICANT: Sklar, Robert
;  APPLICANT: Marchionni, Mark
;  APPLICANT: Gwynne, David I.
;  TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
;  TITLE OF INVENTION: DISORDERS
;  FILE REFERENCE: 04585/028003
;  CURRENT APPLICATION NUMBER: US/08/467,602C
;  CURRENT FILING DATE: 1995-06-06
;  EARLIER APPLICATION NUMBER: 08/209,204
;  EARLIER FILING DATE: 1994-03-08
;  EARLIER APPLICATION NUMBER: 08/059,022
;  EARLIER FILING DATE: 1993-05-06
;  NUMBER OF SEQ ID NOS: 420
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-08-467-602-134

  Query Match             2.0%;  Score 29;  DB 4;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 14
PCT-US94-05083C-130
; Sequence 130, Application PC/TUS9405083C
;  GENERAL INFORMATION:
;    APPLICANT:  Robert Sklar, Mark Marchionni,
;    APPLICANT:  David I. Gwynne
;    TITLE OF INVENTION:  METHODS FOR ALTERING
;    TITLE OF INVENTION:  MUSCLE CONDITION
;    NUMBER OF SEQUENCES:  185
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360
;      MEDIUM TYPE:  kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05083C
;      FILING DATE:  06-MAY-94
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/209,204
;      FILING DATE:  08-MAR-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/059,022
;      FILING DATE:  06-May-93
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/028WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  130:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1193
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US94-05083C-130

  Query Match             2.0%;  Score 29;  DB 5;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


RESULT 15
PCT-US95-06846A-134
```

```
; Sequence 134, Application PC/TUS9506846A
;  GENERAL INFORMATION:
;     APPLICANT:  Goodearl, Andrew David; Stroobant, Paul;
;     APPLICANT:  Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
;     APPLICANT:  Chen, Maio Su; Hiles, Ian
;     TITLE OF INVENTION:  Glial Mitogenic Factors, Their
;     TITLE OF INVENTION:  Preparation and Use
;     NUMBER OF SEQUENCES:  178
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Felfe & Lynch
;       STREET:  805 Third Avenue
;       CITY:  New York City
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;       COMPUTER:  IBM
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06846A
;       FILING DATE:  25-MAY-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/249,322
;       FILING DATE:  26-MAY-1994
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/036,555
;       FILING DATE:  24-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/965,173
;       FILING DATE:  23-OCT-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/940,389
;       FILING DATE:  03-SEP-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/907,138
;       FILING DATE:  30-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/863,703
;       FILING DATE:  03-APRIL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.K. 91 07566.3
;       FILING DATE:  10-APRIL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, Norman D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5250.5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  134:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1193
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US95-06846A-134

  Query Match             2.0%;  Score 29;  DB 5;  Length 1193;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          |||||||||||||||||||||||||||||
Db   1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1152


Search completed: November 30, 2002, 14:02:14
Job time : 69 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:43:44 ; Search time 64 Seconds
(without alignments)
8520.679 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 192557720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1160	81.9	1382	10	US-09-925-300-692	Sequence 692, App
c	2	326	23.0	377	9	US-09-736-457-1096	Sequence 1096, Ap
c	3	326	23.0	377	9	US-09-902-941-1096	Sequence 1096, Ap
c	4	87	6.1	2261	10	US-09-920-300A-1684	Sequence 1684, Ap
c	5	87	6.1	2261	12	US-10-033-528-1684	Sequence 1684, Ap
	6	32	2.3	206	10	US-09-783-590-11058	Sequence 11058, A
c	7	29	2.0	311	10	US-09-867-701-5968	Sequence 5968, Ap
c	8	29	2.0	564	10	US-09-880-107-256	Sequence 256, App
	9	28	2.0	280	10	US-09-960-352-12666	Sequence 12666, A
c	10	28	2.0	319	10	US-09-867-701-2350	Sequence 2350, Ap
	11	28	2.0	941	10	US-09-770-445-359	Sequence 359, App
	12	28	2.0	1505	10	US-09-925-301-142	Sequence 142, App
	13	28	2.0	3238	10	US-09-874-628-5	Sequence 5, Appli
	14	28	2.0	3592	10	US-09-777-745-1	Sequence 1, Appli
	15	28	2.0	3627	10	US-09-777-745-5	Sequence 5, Appli
	16	28	2.0	7878	10	US-09-962-436-562	Sequence 562, App
	17	27	1.9	426	10	US-09-777-564-451	Sequence 451, App
	18	27	1.9	639	9	US-09-950-933A-12	Sequence 12, Appl
	19	27	1.9	915	10	US-09-925-301-244	Sequence 244, App
	20	27	1.9	989	10	US-09-272-162-1	Sequence 1, Appli
	21	27	1.9	1169	10	US-09-804-156-10	Sequence 10, Appl
	22	27	1.9	1334	10	US-09-789-561-30	Sequence 30, Appl
	23	27	1.9	1575	10	US-09-833-381-1159	Sequence 1159, Ap
c	24	27	1.9	1689	9	US-09-989-919-71	Sequence 71, Appl
	25	27	1.9	1969	10	US-09-070-844-7	Sequence 7, Appli
	26	27	1.9	2096	10	US-09-070-844-19	Sequence 19, Appl
	27	27	1.9	2099	10	US-09-070-844-3	Sequence 3, Appli
	28	27	1.9	2137	10	US-09-070-844-18	Sequence 18, Appl
	29	27	1.9	2140	10	US-09-070-844-1	Sequence 1, Appli
	30	27	1.9	3441	9	US-09-944-413-6	Sequence 6, Appli
	31	27	1.9	3441	9	US-09-944-403-6	Sequence 6, Appli
	32	27	1.9	3441	9	US-09-944-896-6	Sequence 6, Appli
	33	27	1.9	3441	9	US-09-944-944-6	Sequence 6, Appli
	34	27	1.9	3441	10	US-09-866-028-6	Sequence 6, Appli
	35	27	1.9	3441	10	US-09-944-449-6	Sequence 6, Appli
	36	27	1.9	3441	10	US-09-944-457-6	Sequence 6, Appli
	37	27	1.9	3441	10	US-09-945-587-6	Sequence 6, Appli
	38	27	1.9	3441	10	US-09-945-015-6	Sequence 6, Appli
	39	27	1.9	3441	10	US-09-944-396-6	Sequence 6, Appli
	40	27	1.9	3441	10	US-09-944-097-6	Sequence 6, Appli
	41	27	1.9	3441	10	US-09-944-432-6	Sequence 6, Appli
	42	27	1.9	3441	10	US-09-943-762-6	Sequence 6, Appli
	43	27	1:9	3441	10	US-09-944-654-6	Sequence 6, Appli
	44	27	1.9	3441	10	US-09-943-851A-6	Sequence 6, Appli
	45	27	1.9	3547	9	US-10-044-716-7	GENERAL INFORMATI

ALIGNMENTS

RESULT 1
US-09-925-300-692
; Sequence 692, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
; LENGTH: 1382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-692

Query Match 81.9%; Score 1160; DB 10; Length 1382;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 138

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 198

Qy    230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 258

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 318
```

```
Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 378

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 438

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 498

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 558

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    559 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGG-CTATGTGTGAAAGCAACTCTAC 617

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 677

Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 737

Qy    770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 797

Qy    830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 857

Qy    890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 917

Qy    950 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 1009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 977

Qy   1010 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAA 1037

Qy   1070 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 TCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACC 1097

Qy   1130 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 CCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCT 1157

Qy   1190 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 CCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGA 1217

Qy   1250 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 AACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATA 1277

Qy   1310 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 TAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1337

Qy   1370 GAAAAATATAATAATCATAAA 1390
          ||||||||||||||||||||
Db   1338 GAAAAATATAATAATCATAAA 1358
```

```
RESULT 2
US-09-736-457-1096/c
; Sequence 1096, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-736-457-1096

  Query Match             23.0%;  Score 326;  DB 9;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.3e-121;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy    189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy    249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy    309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy    369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy    429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1
```

```
RESULT 3
US-09-902-941-1096/c
; Sequence 1096, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
```

```
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-902-941-1096

  Query Match             23.0%;  Score 326;  DB 9;  Length 377;
  Best Local Similarity   99.7%;  Pred. No. 1.3e-121;
  Matches  376;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    129 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy    189 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy    249 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy    309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 368
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 GCTGGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCA 138

Qy    369 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 TTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGG 78

Qy    429 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 18

Qy    489 GGTTTCCATGAGCGTGG 505
          |||||||||||||||||
Db     17 GGTTTCCATGAGCGTGG 1


RESULT 4
US-09-920-300A-1684/c
; Sequence 1684, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547
;  CURRENT APPLICATION NUMBER: US/09/920,300A
;  CURRENT FILING DATE:  2001-07-31
;  NUMBER OF SEQ ID NOS: 1789
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1684
;   LENGTH: 2261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-920-300A-1684

  Query Match              6.1%;  Score 87;  DB 10;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 5
US-10-033-528-1684/c
; Sequence 1684, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547C1
;  CURRENT APPLICATION NUMBER: US/10/033,528
;  CURRENT FILING DATE:  2001-12-26
;  NUMBER OF SEQ ID NOS: 1896
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1684
;   LENGTH: 2261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-033-528-1684

  Query Match              6.1%;  Score 87;  DB 12;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-26;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1307 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202

Qy   1367 ATTGAAAAATATAATAATCATAAAGTC 1393
          |||||||||||||||||||||||||||
Db   2201 ATTGAAAAATATAATAATCATAAAGTC 2175


RESULT 6
US-09-783-590-11058
; Sequence 11058, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11058
;   LENGTH: 206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: misc feature
;   LOCATION: (5)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (47)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (52)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (60)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (76)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (149)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (164)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (172)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (187)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (190)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (198)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (204)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-11058

  Query Match             2.3%;  Score 32;  DB 10;  Length 206;
  Best Local Similarity   100.0%;  Pred. No. 0.00064;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 GTAGGACCTACTCCCTATCCCGTCGGCCGCGG 110
           ||||||||||||||||||||||||||||||||
Db       8 GTAGGACCTACTCCCTATCCCGTCGGCCGCGG 39


RESULT 7
US-09-867-701-5968/c
; Sequence 5968, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5968
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-5968

  Query Match             2.0%;  Score 29;  DB 10;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 0.0098;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db     40 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 12


RESULT 8
US-09-880-107-256/c
; Sequence 256, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 256
;   LENGTH: 564
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA127741
US-09-880-107-256

  Query Match             2.0%;  Score 29;  DB 10;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 0.0093;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           |||||||||||||||||||||||||||||
Db     43 AAAGTCAAAAAAAAAAAAAAAAAAAAAAA 15


RESULT 9
US-09-960-352-12666
; Sequence 12666, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12666
;   LENGTH: 280
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 54-LIB3058-028-Q1-K1-F6
US-09-960-352-12666

  Query Match             2.0%;  Score 28;  DB 10;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
           ||||||||||||||||||||||||||||
Db    200 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 227


RESULT 10
US-09-867-701-2350/c
```

```
; Sequence 2350, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2350
;   LENGTH: 319
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(319)
;   OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2350

  Query Match             2.0%;  Score 28;  DB 10;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db     43 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 16


RESULT 11
US-09-770-445-359
; Sequence 359, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
;  APPLICANT:  Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Rameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION:  Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE:  2023US (PARA-012PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,445
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER:  US 60/178,472
;  PRIOR FILING DATE:  2000-01-27
;  NUMBER OF SEQ ID NOS:  999
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 359
;   LENGTH: 941
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(941)
;   OTHER INFORMATION: n = A,T,C or G
US-09-770-445-359

  Query Match             2.0%;  Score 28;  DB 10;  Length 941;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db    912 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 939


RESULT 12
US-09-925-301-142
; Sequence 142, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
;   LENGTH: 1505
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1493)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1499)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1500)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-142

  Query Match             2.0%;  Score 28;  DB 10;  Length 1505;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   1432 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1459


RESULT 13
US-09-874-628-5
; Sequence 5, Application US/09874628
; Patent No. US20020137133A1
;   GENERAL INFORMATION:
;        APPLICANT: WOZNEY, John
;                   CELESTE, Anthony J.
;                   THIES, R. Scott
;                   YAMAJI, No. US20020137133A1oru
;        TITLE OF INVENTION: RECEPTOR PROTEINS
;        NUMBER OF SEQUENCES: 19
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genetics Institute Inc.- Legal Affairs
;             STREET: 87 CambridgePark Drive
;             CITY: Cambridge
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02140
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.25
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/874,628
;              FILING DATE: 05-Jun-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/123,934
;              FILING DATE: 17-SEP-1993
;         ATTORNEY/AGENT INFORMATION:
;              NAME: LAZAR, Steven R
;              REGISTRATION NUMBER: 32,618
;              REFERENCE/DOCKET NUMBER: 5203
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 617 876 1170
;              TELEFAX: 617 876 5851
;    INFORMATION FOR SEQ ID NO: 5:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 3238 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         IMMEDIATE SOURCE:
;              CLONE: CFK1-10a
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  474..2000
;         SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-874-628-5

  Query Match             2.0%;  Score 28;  DB 10;  Length 3238;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3155 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3182


RESULT 14
US-09-777-745-1
; Sequence 1, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
;  APPLICANT:  CALMELS, THIERRY PAUL GERARD
;  APPLICANT:  SOUCHET, MICHEL LOUIS
;  APPLICANT:  LEGER, ISABELLE MARIE
;  APPLICANT:  TOURTELIER, LAURENCE NATHALIE PATRICIA
;  APPLICANT:  BRIL, ANTOINE MICHEL ALAIN
;  APPLICANT:  DOE, TRUDY RACHEL
;  TITLE OF INVENTION:  NOVEL COMPOUNDS
;  FILE REFERENCE:  GH-30321-C1
;  CURRENT APPLICATION NUMBER: US/09/777,745
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER:  EP 98400014.1
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  EP 98400015.8
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  UK 9710910.2
;  PRIOR FILING DATE:  1997-05-27
;  PRIOR APPLICATION NUMBER:  09/082,271
;  PRIOR FILING DATE:  1998-05-20
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 3592
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-777-745-1

  Query Match             2.0%;  Score 28;  DB 10;  Length 3592;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3559 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3586


RESULT 15
US-09-777-745-5
; Sequence 5, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
;  APPLICANT:  CALMELS, THIERRY PAUL GERARD
;  APPLICANT:  SOUCHET, MICHEL LOUIS
;  APPLICANT:  LEGER, ISABELLE MARIE
;  APPLICANT:  TOURTELIER, LAURENCE NATHALIE PATRICIA
;  APPLICANT:  BRIL, ANTOINE MICHEL ALAIN
;  APPLICANT:  DOE, TRUDY RACHEL
;  TITLE OF INVENTION:  NOVEL COMPOUNDS
;  FILE REFERENCE:  GH-30321-C1
;  CURRENT APPLICATION NUMBER: US/09/777,745
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER:  EP 98400014.1
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  EP 98400015.8
;  PRIOR FILING DATE:  1998-01-07
;  PRIOR APPLICATION NUMBER:  UK 9710910.2
;  PRIOR FILING DATE:  1997-05-27
;  PRIOR APPLICATION NUMBER:  09/082,271
;  PRIOR FILING DATE:  1998-05-20
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3627
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-777-745-5

  Query Match             2.0%;  Score 28;  DB 10;  Length 3627;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1389 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 1416
          ||||||||||||||||||||||||||||
Db   3594 AAGTCAAAAAAAAAAAAAAAAAAAAAAA 3621


Search completed: November 30, 2002, 14:03:47
Job time : 83 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:38:19 ; Search time 1858 Seconds
(without alignments)
12342.739 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gcccactcgctgcggcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	844	59.6	844	9	AL529152	AL529152 AL529152
	2	813	57.4	890	9	AL520768	AL520768 AL520768
	3	798	56.4	943	14	BM804704	BM804704 AGENCOURT
c	4	691	48.8	960	9	AL522373	AL522373 AL522373
	5	677	47.8	782	14	BQ441752	BQ441752 AGENCOURT
	6	675	47.7	1048	14	BQ059007	BQ059007 AGENCOURT
c	7	660	46.6	677	14	BQ581760	BQ581760 il12g06.x
c	8	645	45.6	659	13	BM666784	BM666784 UI-E-CL1-
c	9	625	44.1	625	13	BI711300	BI711300 id96f12.x
c	10	610	43.1	729	14	BM975958	BM975958 UI-CF-EN1
	11	606	42.8	613	14	BM690873	BM690873 UI-E-CI1-
	12	604	42.7	846	13	BI769604	BI769604 603054985
c	13	589	41.6	750	14	BM979227	BM979227 UI-CF-DU1
c	14	585	41.3	992	9	AL520767	AL520767 AL520767
c	15	584	41.2	586	9	AI797259	AI797259 we86e02.x
	16	577	40.7	600	14	BM726885	BM726885 UI-E-EJ0-
	17	573	40.5	1150	14	BM925975	BM925975 AGENCOURT
	18	572	40.4	579	13	BI711712	BI711712 id96f12.y
c	19	571	40.3	736	9	AI668782	AI668782 wc14b12.x
	20	563	39.8	624	14	BQ582098	BQ582098 il12g06.y
c	21	557	39.3	559	13	BM141822	BM141822 if26h04.x
c	22	548	38.7	550	13	BM141736	BM141736 if25d10.x
c	23	547	38.6	766	14	BM679998	BM679998 UI-E-EO1-
	24	545	38.5	793	12	BG436916	BG436916 602488424
	25	543	38.3	746	13	BI752623	BI752623 603028471
	26	539	38.1	938	14	BQ073413	BQ073413 AGENCOURT
	27	534	37.7	764	12	BG762506	BG762506 602733954
c	28	524	37.0	550	9	AI017543	AI017543 ou35c05.x
c	29	510	36.0	710	9	AI797380	AI797380 we87d12.x
	30	510	36.0	781	12	BG532265	BG532265 602561355
	31	510	36.0	928	13	BI822844	BI822844 603040166
	32	505	35.7	537	13	BM142010	BM142010 if25d10.y
	33	503	35.5	552	13	BM142111	BM142111 if26h04.y
c	34	487	34.4	560	10	AW182514	AW182514 xj81b03.x
	35	468	33.1	576	10	AW956706	AW956706 EST368776
	36	468	33.1	657	13	BI916460	BI916460 603182527
c	37	461	32.6	461	9	AA907376	AA907376 oj70h11.s
	38	459	32.4	946	13	BI757823	BI757823 603030404
	39	457	32.3	790	12	BG703176	BG703176 602686082
	40	452	31.9	813	13	BI755493	BI755493 603027349
c	41	451	31.9	633	14	BQ002524	BQ002524 UI-H-EI1-
	42	448	31.6	791	12	BG776560	BG776560 602663693
c	43	446	31.5	579	9	AI559188	AI559188 tq42g05.x
c	44	444	31.4	749	10	AW073366	AW073366 xa39c01.x
c	45	434	30.6	482	9	AI085503	AI085503 qf20c07.x

ALIGNMENTS

RESULT 1
AL529152
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5 prime, mRNA sequence.
ACCESSION AL529152
VERSION AL529152.1 GI:12792645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YM06"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

```
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      182 a    238 c    235 g    189 t
ORIGIN

  Query Match           59.6%;  Score 844;  DB 9;  Length 844;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  844;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    300 AAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTG 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTG 60

Qy    360 CCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCA 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCA 120

Qy    420 CTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 180

Qy    480 TCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAAT 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAAT 240

Qy    540 TGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTG 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTG 300

Qy    600 TGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGG 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGG 360

Qy    660 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTAT 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTAT 420

Qy    720 GACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACC 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACC 480

Qy    780 TATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCC 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCC 540

Qy    840 CGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAG 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAG 600

Qy    900 AAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGC 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGC 660

Qy    960 TGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTG 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTG 720

Qy   1020 CGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCAC 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCAC 780

Qy   1080 CCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGC 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGC 840

Qy   1140 CACT 1143
          ||||
Db    841 CACT 844


RESULT 2
AL520768
LOCUS       AL520768                 890 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
            prime, mRNA sequence.
ACCESSION   AL520768
VERSION     AL520768.1  GI:12784261
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      191 a    251 c    241 g    206 t      1 others
ORIGIN

  Query Match           57.4%;  Score 813;  DB 9;  Length 890;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  813;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    118 TCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTC 177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTC 129

Qy    178 AACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCT 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 AACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCT 189

Qy    238 GCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACT 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 GCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACT 249

Qy    298 TTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 TTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 309

Qy    358 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAAC 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAAC 369

Qy    418 CACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGC 477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    370 CACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGC 429

Qy    478 TGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACA 537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACA 489

Qy    538 ATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATC 597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 ATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATC 549

Qy    598 TGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 TGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 609

Qy    658 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCT 717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCT 669

Qy    718 ATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTA 777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 ATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTA 729

Qy    778 CCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGG 837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 CCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGG 789

Qy    838 CCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATG 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 CCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATG 849

Qy    898 AGAAGAGAGCAAGTTATGGCCACAGCATGGTGG 930
          ||||||||||||||||||||||||||||||||
Db    850 AGAAGAGAGCAAGTTATGGCCACAGCATGGTGG 882
```

RESULT 3
BM804704
LOCUS BM804704 943 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6494868 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725375 5', mRNA sequence.
ACCESSION BM804704
VERSION BM804704.1 GI:19121527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 943)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12716 row: c column: 08
High quality sequence stop: 716.
FEATURES Location/Qualifiers
source 1. .943
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5725375"
/clone_lib="NIH_MGC_125"
/lab_host="DH10B"
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."
BASE COUNT 198 a 270 c 260 g 215 t
ORIGIN

Query Match 56.4%; Score 798; DB 14; Length 943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 113

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 173

Qy    230 TTCCTCCTGCGAACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 TTCCTCCTGCGAACTGCCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 233

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 293

Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    294 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 353

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 413

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    414 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 473

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 533

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 593

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 653

Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 713

Qy    770 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    714 GATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTT 773

Qy    830 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    774 GCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACG 833

Qy    890 CCACCATGAGAAGAGAGC 907
          ||||||||||||||||||
Db    834 CCACCATGAGAAGAGAGC 851
```

RESULT 4
AL522373/c
LOCUS AL522373 960 bp mRNA linear EST 13-FEB-2001
DEFINITION AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3

```
          prime, mRNA sequence.
ACCESSION   AL522373
VERSION     AL522373.1  GI:12785866
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB008YM02"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      225 a    259 c    255 g    217 t      4 others
ORIGIN
```

Query Match 48.8%; Score 691; DB 9; Length 960;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 911; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy    438 GAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCAT 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 GAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCAT 863

Qy    498 GAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 GAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 803

Qy    558 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCA 617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCA 743

Qy    618 GGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCT 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 GGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCT 683

Qy    678 GTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAA 737
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    682 GTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGC-GTTCCCTGAA 624

Qy    738 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGA 797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    623 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGA 564

Qy    798 TCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 TCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 504

Qy    858 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGC 917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    503 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGC 444
```

```
Qy    918 CACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGC 977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    443 CACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGC 384

Qy    978 CTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTG 1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    383 CTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTG 324

Qy   1038 TTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTG 1097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    323 TTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTG 264

Qy   1098 ACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTC 1157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    263 ACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTC 204

Qy   1158 ATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCT 1217
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    203 ATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCMCTTGGAGAACCTTGACTCTCT 144

Qy   1218 TGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTC 1277
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    143 TGATGGAACACAGATGGGCTGCTTGGGAAAGAMACTTTCACCTGAGCTTCACCTGAGGTC 84

Qy   1278 AGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGA 1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83 AGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGA 24

Qy   1338 AGTTCTGCTGAGGG 1351
          ||||||||||||||
Db     23 AGTTCTGCTGAGGG 10
```

```
RESULT 5
BQ441752
LOCUS       BQ441752                782 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7761426 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6019908
            5', mRNA sequence.
ACCESSION   BQ441752
VERSION     BQ441752.1  GI:21180828
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13223  row: c  column: 13
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .782
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6019908"
                     /clone_lib="NIH_MGC_70"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
```

```
BASE COUNT      185 a    209 c    200 g    186 t      2 others
ORIGIN

  Query Match            47.8%;  Score 677;  DB 14;  Length 782;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  677;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    696 AAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCT 755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCT 60

Qy    756 CAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCC 815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCC 120

Qy    816 CACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCA 875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCA 180

Qy    876 GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGAC 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGAC 240

Qy    936 CCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATA 995
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATA 300

Qy    996 GACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCT 1055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCT 360

Qy   1056 GACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC 1115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC 420

Qy   1116 CTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGG 1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGG 480

Qy   1176 ATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 1235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 540

Qy   1236 CTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAA 1295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAA 600

Qy   1296 GGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGA 1355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGA 660

Qy   1356 GCAGCACTGGCATTGAA 1372
          |||||||||||||||||
Db    661 GCAGCACTGGCATTGAA 677


RESULT 6
BQ059007
LOCUS       BQ059007                1048 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6808192 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5814981
            5', mRNA sequence.
ACCESSION   BQ059007
VERSION     BQ059007.1  GI:19818347
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1048)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2069  row: h  column: 22
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5814981"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      216 a    304 c    289 g    239 t
ORIGIN

  Query Match            47.7%;  Score 675;  DB 14;  Length 1048;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  675;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 124

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 184

Qy    230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 244

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 304

Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    305 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 364

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    365 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 424

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    425 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 484

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    485 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 544

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    545 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 604

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 664
```

```
Qy    710 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    665 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 724

Qy    770 GATACTTACCTATCC 784
          |||||||||||||||
Db    725 GATACTTACCTATCC 739


RESULT 7
BQ581760/c
LOCUS       BQ581760                677 bp    mRNA    linear   EST 20-JUN-2002
DEFINITION  il12g06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029771
            3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1. ;, mRNA
            sequence.
ACCESSION   BQ581760
VERSION     BQ581760.1  GI:21494649
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Sceare,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6029771"
                     /clone_lib="Human insulinoma"
                     /tissue_type="insulinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
                     (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
                     pBluescript SK- by Dr. H. Inoue following the Washington
                     University protocol
                     (http://genome.wustl.edu/est/lambda_protocol.shtml).
                     Please contact Hiroshi Inoue, MD/PhD for further
                     information on this library (Metabolism Division, Permutt
                     Laboratory, Washington University School of Medicine, Box
                     8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
                     is a Washington University Pancreas EST project library. "
BASE COUNT      158 a    170 c    179 g    170 t
ORIGIN

  Query Match             46.6%;  Score 660;  DB 14;  Length 677;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  660;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    732 CCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 CCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 618

Qy    792 TTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAA 851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 TTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAA 558

Qy    852 ACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGT 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGT 498

Qy    912 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGG 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGG 438

Qy    972 CCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTG 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 CCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTG 378

Qy   1032 CCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAA 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 CCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAA 318

Qy   1092 GACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTA 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTA 258

Qy   1152 GTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGA 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 GTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGA 198

Qy   1212 CTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCT 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 CTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCT 138

Qy   1272 GAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGA 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 GAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGA 78

Qy   1332 AACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 AACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 18


RESULT 8
BM666784/c
LOCUS       BM666784                659 bp    mRNA    linear   EST 27-FEB-2002
DEFINITION  UI-E-CL1-aez-a-24-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
            UI-E-CL1-aez-a-24-0-UI 3', mRNA sequence.
ACCESSION   BM666784
VERSION     BM666784.1  GI:18974615
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
```

```
          Clone Distribution: Researchers may obtain clones from Research
          Genetics (www.resgen.com).
          The following repetitive elements were found in this cDNA
          sequence: 1-35, >AT_rich#Low_complexity
          Seq primer: M13 Forward
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL1-aez-a-24-0-UI"
                     /clone_lib="UI-E-CL1"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CL1 is a normalized cDNA library containing the
                     following tissue(s): retina. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CCGCG. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI).
                     TAG_LIB=UI-E-CL1
                     TAG_TISSUE=human retina
                     TAG_SEQ=CCGCG"
BASE COUNT      153 a    168 c    173 g    165 t
ORIGIN

  Query Match             45.6%;  Score 645;  DB 13;  Length 659;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  645;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     747 GCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     659 GCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 600

Qy     807 GGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTA 866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     599 GGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTA 540

Qy     867 GTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATG 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 GTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATG 480

Qy     927 GTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTT 986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 GTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTT 420

Qy     987 GCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCAC 1046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 GCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCAC 360

Qy    1047 CGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTG 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 CGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTG 300

Qy    1107 AGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTT 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 AGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTT 240

Qy    1167 GGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAAC 1226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 GGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAAC 180

Qy    1227 ACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAG 1286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 ACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAG 120

Qy    1287 TTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCT 1346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119 TTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCT 60

Qy    1347 GAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 1391
           |||||||||||||||||||||||||||||||||||||||||||||
Db      59 GAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAG 15


RESULT 9
BI711300/c
LOCUS       BI711300                625 bp    mRNA    linear   EST 11-MAR-2002
DEFINITION  id96f12.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023654
            3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1. ;, mRNA
            sequence.
ACCESSION   BI711300
VERSION     BI711300.1  GI:15686995
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Sceare,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5023654"
                     /clone_lib="Human insulinoma"
                     /tissue_type="insulinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
                     (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
                     pBluescript SK- by Dr. H. Inoue  following the Washington
                     University protocol
                     (http://genome.wustl.edu/est/lambda_protocol.shtml).
                     Please contact  Hiroshi Inoue, MD/PhD for further
                     information on this library  (Metabolism Division, Permutt
                     Laboratory, Washington University School  of Medicine, Box
                     8127, 660 S Euclid Ave, St. Louis, MO 63110).  Note: this
                     is a Washington University Pancreas EST project library. "
BASE COUNT      139 a    183 c    161 g    142 t
ORIGIN

  Query Match             44.1%;  Score 625;  DB 13;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  625;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     333 AGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCT 392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     625 AGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCT 566

Qy     393 GCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAG 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     565 GCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAG 506

Qy     453 CTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGAC 512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     505 CTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGAC 446

Qy     513 TGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTA 572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     445 TGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTA 386

Qy     573 GTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 632
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     385 GTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 326

Qy     633 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCA 692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCA 266

Qy     693 GGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTG 752
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 GGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTG 206

Qy     753 GCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCA 812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 GCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCA 146

Qy     813 GCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCA 872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 GCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCA 86

Qy     873 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTA 932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 GCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTA 26

Qy     933 GACCCCTGGGGAACAGTGGTGGCCC 957
           |||||||||||||||||||||||||
Db      25 GACCCCTGGGGAACAGTGGTGGCCC 1
```

```
RESULT 10
BM975958/c
LOCUS       BM975958                729 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-EN1-acw-l-14-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
            UI-CF-EN1-acw-l-14-0-UI 3', mRNA sequence.
ACCESSION   BM975958
VERSION     BM975958.1  GI:19593549
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-35, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EN1-acw-l-14-0-UI"
                     /clone_lib="UI-CF-EN1"
                     /tissue_type="Primary Lung Cystic Fibrosis Epithelial
                     Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EN1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Cystic Fibrosis
                     Epithelial Cells. The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was ligated to an EcoR I adaptor, digested with Not
                     I, and cloned directionally into pT7T3-Pac vector. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CTGCTCAGGT.
                     TAG_LIB=UI-CF-EN1
                     TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
                     6hr to LPS 24h
                     TAG_SEQ=CTGCTCAGGT"
BASE COUNT      171 a    183 c    195 g    179 t      1 others
ORIGIN
```

Query Match 43.1%; Score 610; DB 14; Length 729;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 660; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy     731 CCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGC 790
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     675 CCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGC 616

Qy     791 TTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGA 850
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     615 TTNTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGA 556

Qy     851 AACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     555 AACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 496

Qy     911 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGG 970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     495 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGG 436

Qy     971 GCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCT 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     435 GCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCT 376

Qy    1031 GCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTA 1090
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     375 GCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTA 316

Qy    1091 AGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCT 1150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 AGACTTGACTTCTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCT 256

Qy    1151 AGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTG 1210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    255 AGTGCTCATGTGACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTG 196

Qy   1211 ACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACC 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 ACTCTCTTGATGGAACACAGATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACC 136

Qy   1271 TGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGG 1330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 TGAGGTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGG 76

Qy   1331 AAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAA 1390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 AAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAA 16

Qy   1391 G 1391
          |
Db     15 G 15
```

```
RESULT 11
BM690873
LOCUS       BM690873                 613 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CI1-aaz-a-06-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone
            UI-E-CI1-aaz-a-06-0-UI 5', mRNA sequence.
ACCESSION   BM690873
VERSION     BM690873.1  GI:19004131
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CI1-aaz-a-06-0-UI"
                     /clone_lib="UI-E-CI1"
                     /tissue_type="RPE and Choroid"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CI1 is a normalized cDNA library containing the
                     following tissue(s): RPE and Choroid. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is ACCTA.
                     This library was created for the program, Gene Discovery
                     in the Visual System, supported by National Eye Institute
                     (NEI)."
BASE COUNT      128 a    184 c    164 g    137 t
ORIGIN
```

```
  Query Match           42.8%;  Score 606;  DB 14;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  606;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    645 TCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGT 704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8 TCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGT 67

Qy    705 CTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGA 764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 CTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGA 127

Qy    765 GCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAG 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAG 187

Qy    825 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 247

Qy    885 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 307

Qy    945 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 1004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 367

Qy   1005 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 1064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 427

Qy   1065 GGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTC 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    428 GGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCCTC 487

Qy   1125 CCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCA 1184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 CCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGGCA 547

Qy   1185 CAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGG 1244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 CAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTGGG 607

Qy   1245 AAAGAA 1250
          ||||||
Db    608 AAAGAA 613
```

```
RESULT 12
BI769604
LOCUS       BI769604                 846 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603054985F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
            mRNA sequence.
ACCESSION   BI769604
VERSION     BI769604.1  GI:15761182
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11512  row: j  column: 22
            High quality sequence stop: 838.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5204469"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      179 a    240 c    226 g    201 t
ORIGIN

  Query Match             42.7%;  Score 604;  DB 13;  Length 846;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  604;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    110 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCG 120

Qy    170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTC 180

Qy    230 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCA 240

Qy    290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 300

Qy    350 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCT 360

Qy    410 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGG 420

Qy    470 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAA 480

Qy    530 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAA 540

Qy    590 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 600

Qy    650 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 660

Qy    710 TGTC 713
          ||||
Db    661 TGTC 664


RESULT 13
BM979227/c
LOCUS       BM979227                750 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-DU1-adl-n-04-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
            UI-CF-DU1-adl-n-04-0-UI 3', mRNA sequence.
ACCESSION   BM979227
VERSION     BM979227.1  GI:19599461
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 1-41, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-DU1-adl-n-04-0-UI"
                     /clone_lib="UI-CF-DU1"
                     /tissue_type="Primary Lung Epithelial Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-DU1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Epithelial Cells The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site.  Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GGCTGTAGGC.
                     TAG_LIB=UI-CF-DU1
                     TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
                     TAG_SEQ=GGCTGTAGGC"
BASE COUNT      185 a    171 c    193 g    200 t      1 others
ORIGIN

  Query Match             41.6%;  Score 589;  DB 14;  Length 750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  589;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     823 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     589 AGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 530

Qy     883 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     529 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 470

Qy     943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 1002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     469 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCA 410

Qy    1003 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     409 ACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 350

Qy    1063 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     349 ATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 290

Qy    1123 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     289 TCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGAGGCAGGATCCAGG 230

Qy    1183 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     229 CACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 170

Qy    1243 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 GGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGGTGGAA 110

Qy    1303 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     109 TTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAGGGCTGAGCAGCAC 50

Qy    1363 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAA 1411
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAAAAAAAAAAAAA 1
```

```
RESULT 14
AL520767/c
LOCUS       AL520767                992 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 3
            prime, mRNA sequence.
ACCESSION   AL520767
VERSION     AL520767.1  GI:12784260
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      229 a    267 c    268 g    222 t      6 others
ORIGIN
```

```
  Query Match             41.3%;  Score 585;  DB 9;  Length 992;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches  805;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     443 ATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCG 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 ATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCG 870

Qy     503 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAA 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCAA 810

Qy     563 AGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 AGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 750

Qy     623 GGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 GGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 690

Qy     683 CACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTC 742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 CACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTC 630

Qy     743 TCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCAT 802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 TCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCAT 570

Qy     803 TACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569 TACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 510

Qy     863 TGTAGTGGC-AGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 921
           ||||||| |  ||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 TGTAGTGCCMCGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 450

Qy     922 GCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCT 981
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 GCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCT 390

Qy     982 GCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCC 1041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 GCCTTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCC 330

Qy    1042 AGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTT 1101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 AGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTT 270

Qy    1102 CTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGT 1161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 CTGTGAGTTTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGT 210

Qy    1162 GACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGAT 1221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 GACTTGGAGGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGAT 150

Qy    1222 GGAACACAGATGGGCTGCTTGGGAAAGA 1249
           ||||||||||||||||||||||||||||
Db     149 GGAACACAGATGGGCTGCTTGGGAAAGA 122
```

```
RESULT 15
AI797259/c
```

```
LOCUS       AI797259                586 bp    mRNA    linear   EST 18-DEC-1999
DEFINITION  we86e02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2347994 3' similar to TR:O76091 O76091 NITRILASE HOMOLOG 1.
            ;, mRNA sequence.
ACCESSION   AI797259
VERSION     AI797259.1  GI:5362731
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1500   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2347994"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      137 a    155 c    158 g    136 t
ORIGIN

  Query Match            41.2%;  Score 584;  DB 9;  Length 586;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  584;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     810 CCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTG 869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     586 CCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTG 527

Qy     870 GCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTG 929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     526 GCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTG 467

Qy     930 GTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCC 989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     466 GTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCC 407

Qy     990 CGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGC 1049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     406 CGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGC 347

Qy    1050 AGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGT 1109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 AGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGACTTGACTTCTGTGAGT 287

Qy    1110 TTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGA 1169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     286 TTAGACCTGCCCCTCCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTGGA 227

Qy    1170 GGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACA 1229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     226 GGCAGGATCCAGGCACAGCTCCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACA 167

Qy    1230 GATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTT 1289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     166 GATGGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTCAGACTGCAGTTT 107

Qy    1290 CAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAG 1349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     106 CAGAAAGGTGGAATTTTATATAGTCATTGTTTATTTCATGGAAACTGAAGTTCTGCTGAG 47

Qy    1350 GGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 1393
           |||||||||||||||||||||||||||||||||||||||||||
Db      46 GGCTGAGCAGCACTGGCATTGAAAAATATAATAATCATAAAGTC 3


Search completed: November 30, 2002, 14:00:56
Job time : 1870 secs
```

```
OM of: US-09-357-675C-21 to: GenEmbl:*   out_format : pfs

Date: Apr 29, 2002 11:05 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1h
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091757_3179/app_query.fasta_1.1208
 -DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09357675_@CGN1_1_10065 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 3374.980000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pr:AF069987       + 1748.00 2415.67  3.4e-126  1385  ! AF069987 Homo sapiens nitrilas
gb_ro:AF069988       + 1484.00 2049.81  8.2e-106  1338  ! AF069988 Mus musculus nitrilas
gb_ro:BC021634       + 1476.00 2038.53  3.5e-105  1365  ! BC021634 Mus musculus, nitrila
gb_pr:AF069984       + 1468.50 2018.45  4.6e-104  4079  ! AF069984 Homo sapiens nitrilas
gb_htg:AL354714_4    + 1468.50 1989.30  1.9e-102  110000  ! Continuation (5 of 7) of AL3
gb_htg:AL354714_5    + 1468.50 1989.30  1.9e-102  110000  ! Continuation (6 of 7) of AL3
gb_htg:AL590651      - 1468.50 1985.56  3.1e-102  167863  ! AL590651 Homo sapiens chromc
gb_pr:AL591806       + 1468.50 1983.97  3.8e-102  200822  ! AL591806 Human DNA sequence
gb_htg:AC084821      + 1105.50 1480.02  4.5e-74   211771  ! AC084821 Mus musculus chromc
gb_ro:AF069985       + 1102.50 1509.97  9.6e-76   4481  ! AF069985 Mus musculus nitrilas
gb_htg:AC087229      + 1083.00 1448.75  2.5e-72   213203  ! AC087229 Mus musculus chromc
gb_htg:AC105589      + 1059.00 1417.03  1.4e-70   178418  ! AC105589 Rattus norvegicus c
gb_ov:AF284575       +  968.50 1335.66  4.9e-66   1214  ! AF284575 Xenopus laevis Nit pr
gb_in:AF069989       +  710.50  975.82  5.4e-46   1521  ! AF069989 Drosophila melanogast
gb_htg:AC017755      +  710.50  936.34  8.6e-44   131853  ! AC017755 Drosophila melanoga
gb_in:AC093121       -  710.50  934.18  1.1e-43   168250  ! AC093121 Drosophila melanoga
gb_in:AC093501       -  710.50  933.91  1.2e-43   173494  ! AC093501 Drosophila melanoga
gb_htg:AC006169      +  710.50  933.87  1.2e-43   174279  ! AC006169 Drosophila melanoga
gb_in:AC105293       +  710.50  933.09  1.3e-43   190351  ! AC105293 Drosophila melanoga
gb_in:AE003467       +  710.50  929.10  2.2e-43   298640  ! AE003467 Drosophila melanoga
gb_in:AF069986       +  702.00  964.86  2.2e-45   1385  ! AF069986 Caenorhabditis elegan
gb_pl:AF372904       +  648.50  892.13  2.5e-41   1172  ! AF372904 Arabidopsis thaliana
gb_htg:AC010013      -  632.50  840.14  2.0e-38   34013  ! AC010013 Drosophila melanogas
gb_ba:AE000934       -  537.50  715.54  1.7e-31   15124  ! AE000934 Methanobacterium the
gb_pl:SCYJL126W      +  519.00  711.54  2.9e-31   1307  ! Z49401 S.cerevisiae chromosome
gb_pl:AF284571       +  519.00  709.80  3.6e-31   1591  ! AF284571 Saccharomyces cerevis
gb_pl:SPBC651        +  502.00  661.73  1.7e-28   25350  ! AL035570 S.pombe chromosome I
gb_ba:AE004861       -  500.50  667.48  8.1e-29   10466  ! AE004861 Pseudomonas aeruginc
gb_pr:AF260334       +  493.00  677.68  2.2e-29   1019  ! AF260334 Homo sapiens CUA002 m
gb_pl:SCGCD14        -  493.00  668.77  6.9e-29   2792  ! Z54149 S.cerevisiae GCD14 gene
gb_ba:SYCSLRD        -  490.00  632.03  7.7e-27   110908  ! D64002 Synechocystis sp. PCC
gb_pr:BC020620       +  487.50  670.54  5.5e-29   965  ! BC020620 Homo sapiens, Nit prot
gb_pr:AF284574       +  485.50  667.79  7.8e-29   962  ! AF284574 Homo sapiens Nit prote
gb_ro:AF284573       +  481.00  658.94  2.4e-28   1292  ! AF284573 Mus musculus Nit prot
gb_ba:AB017194       +  479.50  645.32  1.4e-27   4762  ! AB017194 Plectonema boryanum C
gb_ro:BC020153       +  479.00  657.89  2.8e-28   1063  ! BC020153 Mus musculus, RIKEN c
gb_ba:AP003587       +  477.50  604.69  2.6e-25   343550  ! AP003587 Nostoc sp. PCC 7120
gb_ba:AF005512       -  476.50  644.73  1.5e-27   3180  ! AF005512 Nostoc commune tRNA-L
gb_sts:G72919        +  471.00  648.81  8.9e-28   847  ! G72919 MARC 2849-2850:991933517
gb_ba:AL646071       +  468.00  596.16  7.6e-25   203050  ! AL646071 Ralstonia solanacea
gb_ba:SME591791      -  444.50  558.98  9.0e-23   340900  ! AL591791 Sinorhizobium me
gb_ba:AJ414158       +  440.50  556.73  1.2e-22   235050  ! AJ414158 Yersinia pestis
gb_ba:AE004129       +  436.50  578.36  7.5e-24   10881  ! AE004129 Vibrio cholerae c
gb_ba:AE005759       -  433.50  572.80  1.5e-23   12748  ! AE005759 Caulobacter cresc
gb_pl:SPAC26A3       +  430.50  558.94  9.0e-23   38149  ! Z69240 S.pombe chromosome

seq_name: gb_pr:AF069987

seq_documentation_block:
LOCUS       AF069987               1385 bp    mRNA    linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.
ACCESSION   AF069987
VERSION     AF069987.1  GI:3228665
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and fhit homologs are encoded as fusion proteins in
            drosophila melanogaster and caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
REFERENCE   2  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            1. .1385
                     /gene="NIT1"
     CDS             77. .1060
                     /gene="NIT1"
                     /note="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC39907.1"
                     /db_xref="GI:3228666"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
BASE COUNT      322 a    382 c    357 g    324 t
ORIGIN


alignment_scores:
           Quality: 1748.00             Length:    327
             Ratio:  5.346                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x AF069987   ..

 Align seg 1/1  to: AF069987  from: 1  to: 1385

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 126

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 176

 34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
177 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 226

 51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
227 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 276

 67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
277 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 326

 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
327 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 426

117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
427 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 476

134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
477 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 526

151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
527 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 576

167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
577 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 626

184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
627 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 676

201 AlaValCysTyrAspMetArgProProGluLeuSerLeuAlaLeuAlaGl 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
677 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 726

217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
727 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 776

234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
777 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 826

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
827 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 876

267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
877 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 926

284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
927 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 976

301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
977 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1026

317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
    |||||||||||||||||||||||||||||
```

```
    1027 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1057

seq_name: gb_ro:AF069988

seq_documentation_block:
LOCUS       AF069988               1338 bp    mRNA    linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase 1 (Nit1) mRNA, complete cds.
ACCESSION   AF069988
VERSION     AF069988.1  GI:3228667
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and fhit homologs are encoded as fusion proteins in
            drosophila melanogaster and caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
REFERENCE   2  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1338
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            1. .1338
                     /gene="Nit1"
     CDS             58. .1029
                     /gene="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC40185.1"
                     /db_xref="GI:3228668"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      347 a    335 c    330 g    326 t
ORIGIN


alignment_scores:
           Quality: 1484.00            Length:    327
             Ratio:  4.834               Gaps:      2
 Percent Similarity:  93.884   Percent Identity: 84.404

alignment_block:
 US-09-357-675C-21 x AF069988   ..

 Align seg 1/1  to: AF069988  from: 1  to: 1338

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||:::         ||||||||
      58 ATGCTGGGCTTCATCACCAGGCCTCCTCACCAA.........CTCCTGTG 98

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         |   |||   |||:::   :::      |||||||||:::||||||||||
      99 TACCGGATACCGATTACTTCGAACCCCAGTACTTTGTACTCAGCCCAGGC 148

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         |||||:::|||:::   ||||||:::|||   |||||||||||||||||
```

```
   149 CCAGAACCATGTCC...TCATCAACTTCCTGGGAGCTGCCCCTGGTGGCT 195

    51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
       ||||||||||||||||||||||||:::|||||||:::|||||||||||||||
   196 GTGTGCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACATG 245

    67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
       ||||||||||||:::|||||||||||||||||||||||||||||||||||
   246 TGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCCTTTC 295

    84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
       ||||||||||||||||||||||||||||:::|||||||||||||||||
   296 TGCCTGAGGCATTTGACTTTATTGCACGAAACCCTGCCGAGACATTACTC 345

   101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
       ||||||||||||||:::|||   ||||||   :::|||:::|||||||||
   346 CTGTCCGAACCACTGAATGGGGATCTTTTGGGCCAATATAGCCAGCTTGC 395

   117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
       ||||||||||||:::|||||||||||||||||||||||||||||||||||
   396 CAGGGAATGTGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCC 445

   134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
       ||||||||||||||:::|||||||||||||||||||||||||||||||||
   446 AAGACTGGGAGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAAC 495

   151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
       |||||||||:::|||||||||:::|||||||||||||||||||||||||
   496 AGCAAGGGATCAGTAGTGGCCAGTTACAGGAAGACACATCTGTGCGATGT 545

   167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
       ||||||||||||||||||||||   1|||||||||   |||   ||||
   546 AGAGATCCCAGGTCAGGGGCCGATGAGAGAAAGCAACTATACCAAGCCTG 595

   184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
       ||   :::|||||||   |||||||:::|||||||||||||||:::||||||
   596 GAGGCACTCTTGAGCCACCTGTCAAGACACCGGCTGGCAAGGTTGGTCTA 645

   201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
       |||:::|||||||||||||||||||||||||||||||   ||||||||
   646 GCAATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTCA 695

   217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
       ||||||||||||||||||||||||||||||||||||||||||||:::||||
   696 AGCTGGGGCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTACAG 745

   234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
       ||||||||||||||||||||||||||||||||||||||||||||:::|||
   746 GTCCAGCCCACTGGGAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAG 795

   251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
       ||||||||:::|||||||||||||||||||||||||||||   |||||
   796 TGCTATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGC 845

   267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   846 AAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCC 895

   284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
       ||||||||||||||||||||||||||||||||||||||||||||::::::
   896 GCTGCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTT 945

   301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
       |||:::|||:::|||:::||||||||||||||||||||||||||||||||
   946 CTACAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGA 995

   317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
       ||||||||||:::|||||||||||||||||
   996 CCTCTATGGCAGTCTGGGTCATCCACTCTCT 1026

seq_name: gb_ro:BC021634

seq_documentation_block:
```

```
LOCUS       BC021634               1365 bp    mRNA    linear   ROD 22-JAN-2002
DEFINITION  Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
            complete cds.
ACCESSION   BC021634
VERSION     BC021634.1  GI:18204912
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6754855.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Mus musculus"
                     /db_xref="LocusID:27045"
                     /db_xref="taxon:10090"
                     /clone="MGC:13825 IMAGE:4008543"
                     /tissue_type="Mammary tumor metastatized to lung.
                     MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
                     enhancer."
                     /clone_lib="NCI_CGAP_Lu30"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             59. .1030
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAH21634.1"
                     /db_xref="GI:18204913"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYSSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      368 a    334 c    331 g    332 t
ORIGIN


alignment_scores:
           Quality: 1476.00             Length:    327
             Ratio:  4.824                Gaps:      2
 Percent Similarity:  93.578   Percent Identity: 84.098

alignment_block:
 US-09-357-675C-21 x BC021634   ..

 Align seg 1/1  to: BC021634  from: 1  to: 1365

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||:::         ||||||||
```

```
 59 ATGCTGGGCTTCATCACCAGGCCTCCTCACCAA.........CTCCTGTG 99

 17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
100 TACCGGATACCGATTACTTCGAACCCCAGTACTTTGTACTCAGCCCAGGC 149

 34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
150 CCAGAACCATGTCC...TCATCAACTTCCTGGGAGCTGCCCCTGGTGGCT 196

 51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
197 GTGTGCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACATG 246

 67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
247 TGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCCTTTC 296

 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
297 TGCCTGAGGCATTTGACTTTATTGCACGAAACCCTGCCGAGACATTACTC 346

101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
347 CTGTCCGAACCACTGAATGGGGATCTTTTGGGCCAATATAGCCAGCTTGC 396

117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
397 CAGGGAATGTGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCC 446

134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
447 AAGACTGGGAGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAAC 496

151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
497 AGCAAGGGATCAGTAGTGGCCAGTTACAGGAAGACACATCTGTGCGATGT 546

167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
547 AGAGATCCCAGGTCAGGGGCCGATGAGAGAAAGCAACTATACCAAGCCTG 596

184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
597 GAGGCACTCTTGAGCCACCTGTCAAGACACCGGCTGGCAAGGTTGGTCTA 646

201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
647 GCAATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTCA 696

217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
697 AGCTGGGGCAGAAATACTTACTTATTCTTCAGCCTTTGGATCTGTTACAG 746

234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
747 GTCCAGCCCACTGGGAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAG 796

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
797 TGCTATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGC 846

267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
847 AAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCC 896

284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
897 GCTGCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTT 946

301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
947 CTACAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGA 996
```

```
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
997 CCTCTATGGCAGTCTGGGTCATCCACTCTCT 1027

seq_name: gb_pr:AF069984

seq_documentation_block:
LOCUS       AF069984              4079 bp    DNA     linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069984
VERSION     AF069984.1  GI:3242977
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and fhit homologs are encoded as fusion proteins in
            drosophila melanogaster and caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
REFERENCE   2  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     mRNA            join(635. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3569)
                     /gene="NIT1"
                     /note="alternatively spliced"
                     /product="nitrilase homolog 1"
     gene            635. .3569
                     /gene="NIT1"
     exon            635. .680
                     /gene="NIT1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(679. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3267)
                     /gene="NIT1"
                     /note="Nit1; alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC39901.1"
                     /db_xref="GI:3242978"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
     exon            681. .711
                     /gene="NIT1"
                     /note="1A; alternatively spliced"
     exon            712. .741
                     /gene="NIT1"
                     /note="1B; alternatively spliced"
     exon            940. .1053
                     /gene="NIT1"
                     /note="1C; alternatively spliced"
```

```
     exon            1288. .1383
                     /gene="NIT1"
                     /number=2
     exon            1636. .1890
                     /gene="NIT1"
                     /number=3
     exon            2011. .2114
                     /gene="NIT1"
                     /number=4
     exon            2321. .2454
                     /gene="NIT1"
                     /number=5
     exon            2609. .2734
                     /gene="NIT1"
                     /number=6
     exon            3001. .3569
                     /gene="NIT1"
                     /number=7
BASE COUNT      940 a   1082 c   1043 g   1012 t      2 others
ORIGIN


alignment_scores:
            Quality: 1468.50              Length:    660
              Ratio:  4.505                Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x AF069984   ..

 Align seg 1/1  to: AF069984  from: 1  to: 4079

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1289 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 1338

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
         ||||||||||||||||||||||||||||||||||||||||||
    1339 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 1388

      32 ..................................................  32

    1389 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 1438

      32 ..................................................  32

    1439 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 1488

      32 ..................................................  32

    1489 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 1538

      32 ..................................................  32

    1539 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 1588

      33 ..........................................ArgPr 34
                                                   |||||
    1589 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 1638

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1639 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 1688

      51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1689 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 1738

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1739 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 1788

      84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1789 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 1838

     101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1839 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 1888

     117 ..................................................  117

    1889 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 1938

     117 ..................................................  117

    1939 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 1988

     118 .....................ArgGluCysGlyLeuTrpLeuSerLeuGly 127
                               ||||||||||||||||||||||||||||||
    1989 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 2038

     128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2039 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 2088

     144 nCysHisValLeuLeuAsnSer..Lys..................... 152
         ||||||||||||||||||||||  :::
    2089 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 2138

     152 ..................................................  152

    2139 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 2188

     152 ..................................................  152

    2189 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 2238

     152 ..................................................  152

    2239 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 2288

     153 ............................GlyAlaValValAlaThrT 159
                                     ||||||||||||||||||||
    2289 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 2338

     159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2339 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 2388

     176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2389 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 2438

     192 rThrProAlaGlyLysIleGly.Leu..................... 200
         ||||||||||||||||:::||| |||
    2439 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 2488

     200 ..................................................  200

    2489 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 2538

     200 ..................................................  200

    2539 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 2588

     201 ..........................AlaValCysTyrAspMetArg 207
                                   ||||||||||||||||||||
    2589 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 2638

     208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2639 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 2688

     224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    2689 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 2738

     241 .Leu..................................................... 241
          :::
    2739 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 2788

     241 ..................................................... 241

    2789 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 2838

     241 ..................................................... 241

    2839 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 2888

     241 ..................................................... 241

    2889 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 2938

     241 ..................................................... 241

    2939 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 2988

     242 ................LeuArgAlaArgAlaIleGluThrGlnCysTy 252
                         ||||||||||||||||||||||||||||||||
    2989 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 3038

     252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    3039 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 3088

     269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    3089 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 3138

     286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    3139 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 3188

     302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    3189 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 3238

     319 yrGlyAsnLeuGlyHisProLeuSer 327
          |||||||||||||||||||||||||
    3239 ATGGCAATCTGGGTCACCCACTGTCT 3264

seq_name: gb_htg:AL354714_4

seq_documentation_block:
WPCOMMENT
  Sequence split into 7 fragments  LOCUS AL354714 Accession AL354714
     Fragment Name      Begin     End
     AL354714_0             1    110000
     AL354714_1        100001    210000
     AL354714_2        200001    310000
     AL354714_3        300001    410000
     AL354714_4        400001    510000
     AL354714_5        500001    610000
     AL354714_6        600001    699820
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome 1 c


alignment_scores:
            Quality: 1468.50             Length:    660
              Ratio:  4.505               Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x AL354714_4   ..

 Align seg 1/1  to: AL354714_4  from: 1  to: 110000

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18

          ||||||||||||||||||||||||||||||||||||||||||||||||||
  103858 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 103907

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro....... 32
          ||||||||||||||||||||||||||||||||||||||||
  103908 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 103957

      32 ..................................................... 32

  103958 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 104007

      32 ..................................................... 32

  104008 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 104057

      32 ..................................................... 32

  104058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 104107

      32 ..................................................... 32

  104108 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 104157

      33 ..........................................ArgPr 34
                                                   |||||
  104158 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 104207

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104208 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 104257

      51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104258 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 104307

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104308 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 104357

      84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104358 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 104407

     101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104408 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 104457

     117 ..................................................... 117

  104458 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 104507

     117 ..................................................... 117

  104508 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 104557

     118 ....................ArgGluCysGlyLeuTrpLeuSerLeuGly 127
                              ||||||||||||||||||||||||||||||
  104558 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 104607

     128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  104608 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 104657

     144 nCysHisValLeuLeuAsnSer..Lys..................... 152
          ||||||||||||||||||||  :::
  104658 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 104707

     152 ..................................................... 152

  104708 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 104757

     152 ..................................................... 152
```

```
104758 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 104807

   152 ..................................................  152

104808 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 104857

   153 ..............................GlyAlaValValAlaThrT  159
                                     ||||||||||||||||||||
104858 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 104907

   159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
       ||||||||||||||||||||||||||||||||||||||||||||||||||
104908 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 104957

   176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
       ||||||||||||||||||||||||||||||||||||||||||||||||||
104958 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 105007

   192 rThrProAlaGlyLysIleGly.Leu..................... 200
       ||||||||||||||::::||| |||
105008 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 105057

   200 ..................................................  200

105058 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 105107

   200 ..................................................  200

105108 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 105157

   201 ..........................AlaValCysTyrAspMetArg  207
                                ||||||||||||||||||||
105158 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 105207

   208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105208 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 105257

   224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105258 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 105307

   241 .Leu.................................................  241
        :::
105308 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 105357

   241 ..................................................  241

105358 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 105407

   241 ..................................................  241

105408 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 105457

   241 ..................................................  241

105458 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 105507

   241 ..................................................  241

105508 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 105557

   242 ................LeuArgAlaArgAlaIleGluThrGlnCysTy 252
                       ||||||||||||||||||||||||||||||||
105558 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 105607

   252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105608 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 105657

   269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105658 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 105707

   286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105708 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 105757

   302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
       ||||||||||||||||||||||||||||||||||||||||||||||||||
105758 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 105807

   319 yrGlyAsnLeuGlyHisProLeuSer 327
       |||||||||||||||||||||||||
105808 ATGGCAATCTGGGTCACCCACTGTCT 105833

seq_name: gb_htg:AL354714_5

seq_documentation_block:
WPCOMMENT
  Sequence split into 7 fragments  LOCUS AL354714 Accession AL354714
     Fragment Name      Begin     End
     AL354714_0             1    110000
     AL354714_1        100001    210000
     AL354714_2        200001    310000
     AL354714_3        300001    410000
     AL354714_4        400001    510000
     AL354714_5        500001    610000
     AL354714_6        600001    699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome


alignment_scores:
            Quality: 1468.50            Length:    660
              Ratio:  4.505              Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x AL354714_5   ..

 Align seg 1/1  to: AL354714_5  from: 1  to: 110000

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3858 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 3907

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro....... 32
         ||||||||||||||||||||||||||||||||||||||||
    3908 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 3957

      32 ..................................................  32

    3958 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 4007

      32 ..................................................  32

    4008 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 4057

      32 ..................................................  32

    4058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 4107

      32 ..................................................  32

    4108 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 4157

      33 ..........................................ArgPr 34
                                                   |||||
    4158 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 4207

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4208 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 4257

      51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
4258 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 4307

  68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4308 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 4357

  84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4358 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 4407

 101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4408 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 4457

 117 ..................................................... 117

4458 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 4507

 117 ..................................................... 117

4508 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 4557

 118 .....................ArgGluCysGlyLeuTrpLeuSerLeuGly 127
                          ||||||||||||||||||||||||||||||
4558 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 4607

 128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4608 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 4657

 144 nCysHisValLeuLeuAsnSer..Lys..................... 152
     ||||||||||||||||||||  :::
4658 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 4707

 152 ..................................................... 152

4708 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 4757

 152 ..................................................... 152

4758 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 4807

 152 ..................................................... 152

4808 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 4857

 153 ..............................GlyAlaValValAlaThrT 159
                                  ||||||||||||||||||
4858 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 4907

 159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4908 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 4957

 176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4958 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 5007

 192 rThrProAlaGlyLysIleGly.Leu...................... 200
     ||||||||||||||||:::||| |||
5008 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 5057

 200 ..................................................... 200

5058 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 5107

 200 ..................................................... 200

5108 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 5157

 201 ..........................AlaValCysTyrAspMetArg 207
                               ||||||||||||||||||||
5158 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 5207

 208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5208 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 5257

 224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5258 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 5307

 241 .Leu..................................................... 241
      :::
5308 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 5357

 241 ..................................................... 241

5358 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 5407

 241 ..................................................... 241

5408 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 5457

 241 ..................................................... 241

5458 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 5507

 241 ..................................................... 241

5508 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 5557

 242 .................LeuArgAlaArgAlaIleGluThrGlnCysTy 252
                      ||||||||||||||||||||||||||||||||
5558 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 5607

 252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5608 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 5657

 269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5658 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 5707

 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5708 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 5757

 302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5758 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 5807

 319 yrGlyAsnLeuGlyHisProLeuSer 327
     |||||||||||||||||||||||||
5808 ATGGCAATCTGGGTCACCCACTGTCT 5833
```

```
seq_name: gb_htg:AL590651

seq_documentation_block:
LOCUS       AL590651             167863 bp    DNA     linear   HTG 18-AUG-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL590651
VERSION     AL590651.4  GI:13990236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167863)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13751001.
```

```
            ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: bA137A12
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 164729 bases at least Q40
            Consensus quality: 165486 bases at least Q30
            Consensus quality: 166169 bases at least Q20
            Insert size: 167063; sum-of-contigs
            Insert size: 171878; 3.4% error; agarose-fp
            Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
            coverage: 8.11x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9373: contig of 9373 bp in length
            *    9374 9473: gap of      100 bp
            *    9474    29891: contig of 20418 bp in length
            *   29892 29991: gap of      100 bp
            *   29992    36239: contig of 6248 bp in length
            *   36240 36339: gap of      100 bp
            *   36340   103784: contig of 67445 bp in length
            *  103785 103884: gap of      100 bp
            *  103885   114878: contig of 10994 bp in length
            *  114879 114978: gap of      100 bp
            *  114979   126347: contig of 11369 bp in length
            *  126348 126447: gap of      100 bp
            *  126448   137764: contig of 11317 bp in length
            *  137765 137864: gap of      100 bp
            *  137865   163160: contig of 25296 bp in length
            *  163161 163260: gap of      100 bp
            *  163261   167863: contig of 4603 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-137A12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .9373
                     /note="assembly_fragment:02360
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9474. .29891
                     /note="assembly_fragment:01854
                     fragment_chain:1"
     misc_feature    29992. .36239
                     /note="assembly_fragment:00805
                     fragment_chain:1"
     misc_feature    36340. .103784
                     /note="assembly_fragment:01467
                     fragment_chain:1"
     misc_feature    103885. .114878
                     /note="assembly_fragment:00991
                     fragment_chain:2"
     misc_feature    114979. .126347
                     /note="assembly_fragment:01503
                     fragment_chain:2"
     misc_feature    126448. .137764
                     /note="assembly_fragment:01297
                     fragment_chain:2"
     misc_feature    137865. .163160
                     /note="assembly_fragment:00117
                     fragment_chain:2"
     misc_feature    163261. .167863
                     /note="assembly_fragment:01267
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    43927 a  39935 c  40002 g  43197 t    802 others
ORIGIN


alignment_scores:
           Quality: 1468.50              Length:    660
             Ratio:  4.505                 Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x AL590651/rev   ..

 Align seg 1/1  to reverse of: AL590651  from: 1  to: 167863

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  151392 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 151343

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
         ||||||||||||||||||||||||||||||||||||||||||
  151342 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 151293

      32 ..................................................  32

  151292 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 151243

      32 ..................................................  32

  151242 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 151193

      32 ..................................................  32

  151192 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 151143

      32 ..................................................  32

  151142 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 151093

      33 ..............................................ArgPr 34
                                                       |||||
  151092 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 151043

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  151042 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 150993

      51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  150992 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 150943

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  150942 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 150893

      84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  150892 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 150843

     101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  150842 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 150793

     117 ..................................................  117
```

```
150792 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 150743

   117 ..................................................... 117

150742 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 150693

   118 ....................ArgGluCysGlyLeuTrpLeuSerLeuGly 127
                           ||||||||||||||||||||||||||||||
150692 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 150643

   128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150642 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 150593

   144 nCysHisValLeuLeuAsnSer..Lys..................... 152
       ||||||||||||||||||||||  :::
150592 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 150543

   152 ..................................................... 152

150542 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 150493

   152 ..................................................... 152

150492 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 150443

   152 ..................................................... 152

150442 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 150393

   153 ..............................GlyAlaValValAlaThrT 159
                                     ||||||||||||||||||
150392 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 150343

   159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150342 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 150293

   176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150292 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 150243

   192 rThrProAlaGlyLysIleGly.Leu..................... 200
       ||||||||||||||||:::||| |||
150242 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 150193

   200 ..................................................... 200

150192 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 150143

   200 ..................................................... 200

150142 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 150093

   201 ..........................AlaValCysTyrAspMetArg 207
                                 ||||||||||||||||||||
150092 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 150043

   208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150042 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 149993

   224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
       ||||||||||||||||||||||||||||||||||||||||||||||||
149992 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 149943

   241 .Leu..................................................... 241
        :::
149942 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 149893

   241 ..................................................... 241

149892 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 149843

   241 ..................................................... 241

149842 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 149793

   241 ..................................................... 241

149792 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 149743

   241 ..................................................... 241

149742 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 149693

   242 ..................LeuArgAlaArgAlaIleGluThrGlnCysTy 252
                         ||||||||||||||||||||||||||||||||
149692 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 149643

   252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
       ||||||||||||||||||||||||||||||||||||||||||||||||||
149642 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 149593

   269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
       ||||||||||||||||||||||||||||||||||||||||||||||||||
149592 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 149543

   286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
       ||||||||||||||||||||||||||||||||||||||||||||||||||
149542 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 149493

   302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
       ||||||||||||||||||||||||||||||||||||||||||||||||||
149492 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 149443

   319 yrGlyAsnLeuGlyHisProLeuSer 327
       ||||||||||||||||||||||||||
149442 ATGGCAATCTGGGTCACCCACTGTCT 149417

seq_name: gb_pr:AL591806

seq_documentation_block:
LOCUS       AL591806              200822 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
            sequence.
ACCESSION   AL591806
VERSION     AL591806.16  GI:18476709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200822)
  AUTHORS   Harrison,E. .
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17902927.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
```

```
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr1
          RP11-544M22 is from the library RPCI-11.2 constructed by the group
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          This sequence is the entire insert of clone RP11-544M22 The true
          left end of clone RP11-137A12 is at 156538 in this sequence. The
          true right end of clone RP11-381D2 is at 145015 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .200822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-544M22"
                     /clone_lib="RPCI-11.2"
     misc_feature    10395. .10397
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    44098. .44107
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    59585. .59604
                     /note="Sequence from overlapping clone RP11-381D2
                     (AL162592). Assembly confirmed by restriction digest."
     misc_feature    90797. .90883
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    90861. .90883
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    55215 a  49052 c  46183 g  50372 t
ORIGIN


alignment_scores:
           Quality: 1468.50            Length:    660
             Ratio:  4.505              Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x AL591806   ..

 Align seg 1/1  to: AL591806  from: 1  to: 200822

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  172789 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 172838

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
         |||||||||||||||||||||||||||||||||||||||||
  172839 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 172888

      32 ................................................ 32

  172889 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 172938

      32 ................................................ 32

  172939 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 172988

      32 ................................................ 32

  172989 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 173038

      32 ................................................ 32

  173039 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 173088

      33 ..............................................ArgPr 34
                                                       |||||
  173089 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 173138

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173139 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 173188

      51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173189 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 173238

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173239 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 173288

      84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173289 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 173338

     101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173339 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 173388

     117 ................................................ 117

  173389 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 173438

     117 ................................................ 117

  173439 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 173488

     118 ..................ArgGluCysGlyLeuTrpLeuSerLeuGly 127
                             |||||||||||||||||||||||||||||
  173489 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 173538

     128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173539 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 173588

     144 nCysHisValLeuLeuAsnSer..Lys..................... 152
         ||||||||||||||||||||||  :::
  173589 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 173638

     152 ................................................ 152

  173639 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 173688

     152 ................................................ 152

  173689 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 173738

     152 ................................................ 152

  173739 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 173788

     153 ............................GlyAlaValValAlaThrT 159
                                     ||||||||||||||||||
  173789 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 173838

     159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173839 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 173888

     176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  173889 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 173938

     192 rThrProAlaGlyLysIleGly.Leu..................... 200
         ||||||||||||||||:::||| |||
  173939 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 173988
```

```
   200 ..................................................  200

173989 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 174038

   200 ..................................................  200

174039 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 174088

   201 ..........................AlaValCysTyrAspMetArg  207
                                 ||||||||||||||||||||
174089 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 174138

   208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh  224
       ||||||||||||||||||||||||||||||||||||||||||||||||||
174139 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 174188

   224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal.  240
       ||||||||||||||||||||||||||||||||||||||||||||||||
174189 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 174238

   241 .Leu.....................................................  241
        :::
174239 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 174288

   241 ..................................................  241

174289 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 174338

   241 ..................................................  241

174339 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 174388

   241 ..................................................  241

174389 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 174438

   241 ..................................................  241

174439 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 174488

   242 ..................LeuArgAlaArgAlaIleGluThrGlnCysTy  252
                         ||||||||||||||||||||||||||||||||
174489 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 174538

   252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT  269
       ||||||||||||||||||||||||||||||||||||||||||||||||||
174539 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 174588

   269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys  285
       ||||||||||||||||||||||||||||||||||||||||||||||||||
174589 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 174638

   286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr  302
       ||||||||||||||||||||||||||||||||||||||||||||||||||
174639 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 174688

   302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT  319
       ||||||||||||||||||||||||||||||||||||||||||||||||||
174689 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 174738

   319 yrGlyAsnLeuGlyHisProLeuSer  327
       |||||||||||||||||||||||||
174739 ATGGCAATCTGGGTCACCCACTGTCT 174764

seq_name: gb_htg:AC084821

seq_documentation_block:
LOCUS       AC084821              211771 bp    DNA     linear   HTG 29-JAN-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE,
            1 ordered pieces.
ACCESSION   AC084821
VERSION     AC084821.24  GI:18390260
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211771)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-395h6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211771)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jan 29, 2002 this sequence version replaced gi:16076995.
            ------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   211771: contig of 211771 bp in length.
FEATURES              Location/Qualifiers
     source           1. .211771
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /chromosome="1"
                      /clone="rp23-395h6"
                      /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT    55622 a  48688 c  49207 g  58231 t     23 others
ORIGIN


alignment_scores:
           Quality: 1105.50            Length:    676
             Ratio:  3.893              Gaps:      8
 Percent Similarity:  42.012   Percent Identity: 37.722

alignment_block:
 US-09-357-675C-21 x AC084821   ..

 Align seg 1/1  to: AC084821  from: 1  to: 211771

      29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGl 45
         |||:::      |||||||||:::|||:::   ||||||:::|||   ||
     975 TGTGTCATTCACAGGCCCAGAACCATGTCC...TCATCAACTTCCTGGGA 1021

      45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG  62
         ||||||||||||||||||||||||||||||||||||||:::||||||:
    1022 GCTGCCCCTGGTGGCTGTGTGCCAGGTAACATCAACACCAAACAAGCAAG 1071

      62 lnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly  78
         ::|||||||||||||||||||||||||||:::|||||||||||||||||
    1072 AGAACTTTAAAACATGTGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGT 1121

      79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr  95
         ||||||||||||||||||||||||||||||||||||||||||||:::||
    1122 GCTTGCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCACGAAACCC 1171

      95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112
         ||||||||||||   ||||||||||||||:::|||   ||||||   :
    1172 TGCCGAGACATTACTCCTGTCCGAACCACTGAATGGGGATCTTTTGGGCC 1221

     112 luTyrThrGlnLeuAla..............................  117
         ::|||:::|||||||||
```

```
1222 AATATAGCCAGCTTGCCAGGTACAAGGGTAGGGTTTGGGGGGGGGGGGGT 1271

 117 ..................................................  117

1272 TGTGTGTGTAGAACTCCCTGTTGAGCAGTGTTCCTGGATTGCCACAGAGG 1321

 118 .....................................ArgGluCysGlyL  122
                                          ||||||||||||:
1322 GGGTAGAGTTGTGTGTGTCGTGTCCCCCCCCCCCCCCCAGGGAATGTGGAA 1371

 122 euTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGln  138
      ::|||||||||||||||||||||||||||||||||||||||||||||||
1372 TCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCCAAGACTGGGAGCAG 1421

 139 ThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAla..  154
      :::|||||||||||||||||||||||||||||||||||||||||||
1422 AATCAGAAAATCTACAATTGTCATGTGCTTTTGAACAGCAAGGGTGAGAC 1471

 155 Val.................................................  155
      |||
1472 GTCATAGCCCTTAGTTTGCCTCTTCCCAGGCTTCTCTACCTTTCCTCAAC 1521

 155 .................................................  155

1522 TCAGATCTTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTTTTATGC 1571

 155 .................................................  155

1572 TTCTGTTCTCAGCCTACAAGTTACTCTTTGGGAGAAGAGCTAAGCTGGAC 1621

 155 .................................................  155

1622 TTGTAGAGTGTCTGTACTGAGTATTTTCTCTCTCTCTCTCTCTCTCTC 1671

 156 ....................................ValAlaThrTyrA  160
                                         ||||||:::||||
1672 TCTCTCTCTCTCTCTCTCTCTCTCTAGGATCAGTAGTGGCCAGTTACA 1721

 160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys  176
      ||||||||||||||||||||||||||||||||||||||||||||||||
1722 GGAAGACACATCTGTGCGATGTAGAGATCCCAGGTCAGGGGCCGATGAGA 1771

 177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh  193
      |||||||||   |||   ||||||   :::|||||||   ||||||:::||
1772 GAAAGCAACTATACCAAGCCTGGAGGCACTCTTGAGCCACCTGTCAAGAC 1821

 193 rProAlaGly.Lys.....................................  197
      |||||||||| :::
1822 ACCGGCTGGCAAGGTGGGAGTTTGGAAAGGAAATGGGACTATGTTGAAAA 1871

 197 .................................................  197

1872 GTGGTAACACCCCCCATTGGTGCAGGCTAATGAAGGCACTCAGTGTTCTG 1921

 198 .....................................IleGlyL  200
                                           :::||||
1922 TTGAATGGGACAGGGAAGATGAGTGGGTTGTTTCTCATTTCAGGTTGGTC 1971

 200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla  216
      |||||:::|||||||||||||||||||||||||||||||   ||||||
1972 TAGCAATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCT 2021

 217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleTh  233
      |||||||||||||||||||||||||||||||||||||||||||:::||
2022 CAAGCTGGGGCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTAC 2071

 233 rGlyProAlaHisTrp.....................................  238
      ||||||||||||||||
2072 AGGTCCAGCCCACTGGGAGGTAAGAGAATGCCTCTTCATAACATATAGGC 2121

 238 .................................................  238

2122 CTCCTTTAACCTTACTTTTAGTAGTGGGAGCAGAGTGGTACAAACCTAAA 2171

 238 .................................................  238

2172 TCCAAGCACTTCAGTGGATAGCCACAGGTTCAAAGCCAGCATGCCTTGAC 2221

 238 .................................................  238

2222 TAAAAAAGAAAACAAAATAGAAGCTGGGCAGTGGGGGTGCACACCTTTAA 2271

 238 .................................................  238

2272 TCCCAGCACTCAGGAGGCAGAAGCAGGTGGATCTCTGAGTTTGAGGCCAG 2321

 238 .................................................  238

2322 CCTGGTCTACAGAGTGGGTTCCAGGACAGCCAGGGCTACACAGAGAAACC 2371

 238 .................................................  238

2372 CTGTCTCAAAAACAAAAAAGCAACCCACACCTTCAAAGAGTGGGGCATGG 2421

 238 .................................................  238

2422 TGGTGCATGCTCTTCATCTCAGCACTGTGAGGCAAAGGCAGGCAGATCTC 2471

 238 .................................................  238

2472 TGTGGGTCTAAGCCAGCCAGGGTTACACAGGGAGACCTTATCTCAAAAGT 2521

 238 .................................................  238

2522 ACAAAACACAAACACTCCTCCTTTCTACTTGATTTTCTTCCTTTTCATCT 2571

 238 .................................................  238

2572 GGTGAGGAAATTCATTTCTTAGCTAGGTGTGAGTCATGTTAAAGAGTTAG 2621

 238 .................................................  238

2622 TTATTAACTCTAGTTGGTTAGTGAGAAGATTGGGCCCACAGTGCCCGCTT 2671

 238 .................................................  238

2672 CTTGTCACTTGCCTATTAAACGCTGTATGTTCTGGAGTCAACGCACGTCT 2721

 239 .....Glu.ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa  253
           ::: ||||||||||||||||||||||||||:::|||||||||||
2722 CATCCTCAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAGTGCTATGT 2771

 253 lValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrG  270
      |:::|||||||||||||||||||||||||||||   ||||||||||||
2772 AATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGCAAGTTATG 2821

 270 lyHisSerMetValValAspProTrpGlyThrValValAlaArgCysSer  286
      ||||||||||||||||||||||||||||||||||||||||||||||||
2822 GCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGCTCC 2871

 287 GluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGl  303
      |||||||||||||||||||||||||||||||||||:::::|||:::||
2872 GAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTTCTACAACA 2921

 303 nLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrG  320
      |:::|||:::||||||||||||||||||||||||||||||||||||||
2922 GATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGACCTCTATG 2971

 320 lyAsnLeuGlyHisProLeuSer 327
      ||:::|||||||||||||||||
2972 GCAGTCTGGGTCATCCACTCTCT 2994

seq_name: gb_ro:AF069985

seq_documentation_block:
LOCUS       AF069985               4481 bp    DNA     linear   ROD 23-JUL-1998
```

```
DEFINITION  Mus musculus nitrilase homolog 1 (Nit1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069985
VERSION     AF069985.1  GI:3242979
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and fhit homologs are encoded as fusion proteins in
            drosophila melanogaster and caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
REFERENCE   2  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4481
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     mRNA            join(541. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3944)
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /product="nitrilase homolog 1"
     gene            541. .3944
                     /gene="Nit1"
     exon            541. .606
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(605. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3658)
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC40184.1"
                     /db_xref="GI:3242980"
                     /translation="MLGFITRPPHQLLCTGYRLLRIPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
     exon            863. .986
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=2
     exon            1177. .1263
                     /gene="Nit1"
                     /number=3
     exon            1638. .1889
                     /gene="Nit1"
                     /number=4
     exon            2015. .2118
                     /gene="Nit1"
                     /number=5
     exon            2362. .2495
                     /gene="Nit1"
                     /number=6
     exon            2626. .2751
                     /gene="Nit1"
                     /number=7
     exon            3392. .3944
                     /gene="Nit1"
                     /number=8
BASE COUNT     1101 a   1132 c   1133 g   1115 t
ORIGIN


alignment_scores:
            Quality: 1102.50            Length:    680
              Ratio:  3.868              Gaps:      8
 Percent Similarity:  41.912   Percent Identity: 37.353

alignment_block:
 US-09-357-675C-21 x AF069985   ..

 Align seg 1/1  to: AF069985  from: 1  to: 4481

      29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGl 45
         |||:::      |||||||||:::|||:::   ||||||:::|||   ||
    1624 TGTGTCATTCACAGGCCCAGAACCATGTCC...TCATCAACTTCCTGGGA 1670

      45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 62
         ||||||||||||||||||||||||||||||||||||||:::||||||:
    1671 GCTGCCCCTGGTGGCTGTGTGCCAGGTAACATCAACACCAAACAAGCAAG 1720

      62 lnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78
         ::|||||||||||||||||||||||||:::|||||||||||||||||||
    1721 AGAACTTTAAAACATGTGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGT 1770

      79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr 95
         ||||||||||||||||||||||||||||||||||||||||||||:::||
    1771 GCTTGCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCACGAAACCC 1820

      95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112
         ||||||||||||   ||||||||||||||:::|||   ||||||   :
    1821 TGCCGAGACATTACTCCTGTCCGAACCACTGAATGGGGATCTTTTGGGCC 1870

     112 luTyrThrGlnLeuAlaArg..Glu..................... 119
         ::|||:::|||||||||||||  :::
    1871 AATATAGCCAGCTTGCCAGGTACAAGGGTAGGGTTTGGGGGGGGGGGGGT 1920

     119 ..................................................... 119

    1921 TGTGTGTGTAGAACTCCCTGTTGAGCAGTGTTCCTGAATTGCCACAGAGG 1970

     120 .....................................................Cy 120
                                                              ||
    1971 GGGTAGAGTTGTGTGTGTCGTGTCGTGTCCCCCCCCCCCCCCCAGGGAATG 2020

     120 sGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpG 137
         ||||:::|||||||||||||||||||||||||||||||||||||||||||
    2021 TGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCCAAGACTGGG 2070

     137 luGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLys... 152
         |||||:::|||||||||||||||||||||||||||||||||||||||
    2071 AGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAACAGCAAGGGT 2120

     152 ..................................................... 152

    2121 GAGACGTCATAGCCCTTTAGTTTGCCTCTTCCCAGGCTTCTCTACCTTTC 2170

     152 ..................................................... 152

    2171 CTCAACTCAGATCTTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTT 2220

     152 ..................................................... 152

    2221 TTATGCTTTTGTTCTCAGCCTACAAGTTACTCTTTGGGAGAAGAGCTAAG 2270

     152 ..................................................... 152

    2271 CTGGACTTGTAGAGTGTCTGTACTGAGTATTTTCTTTCTGTCTCTCTCTC 2320
```

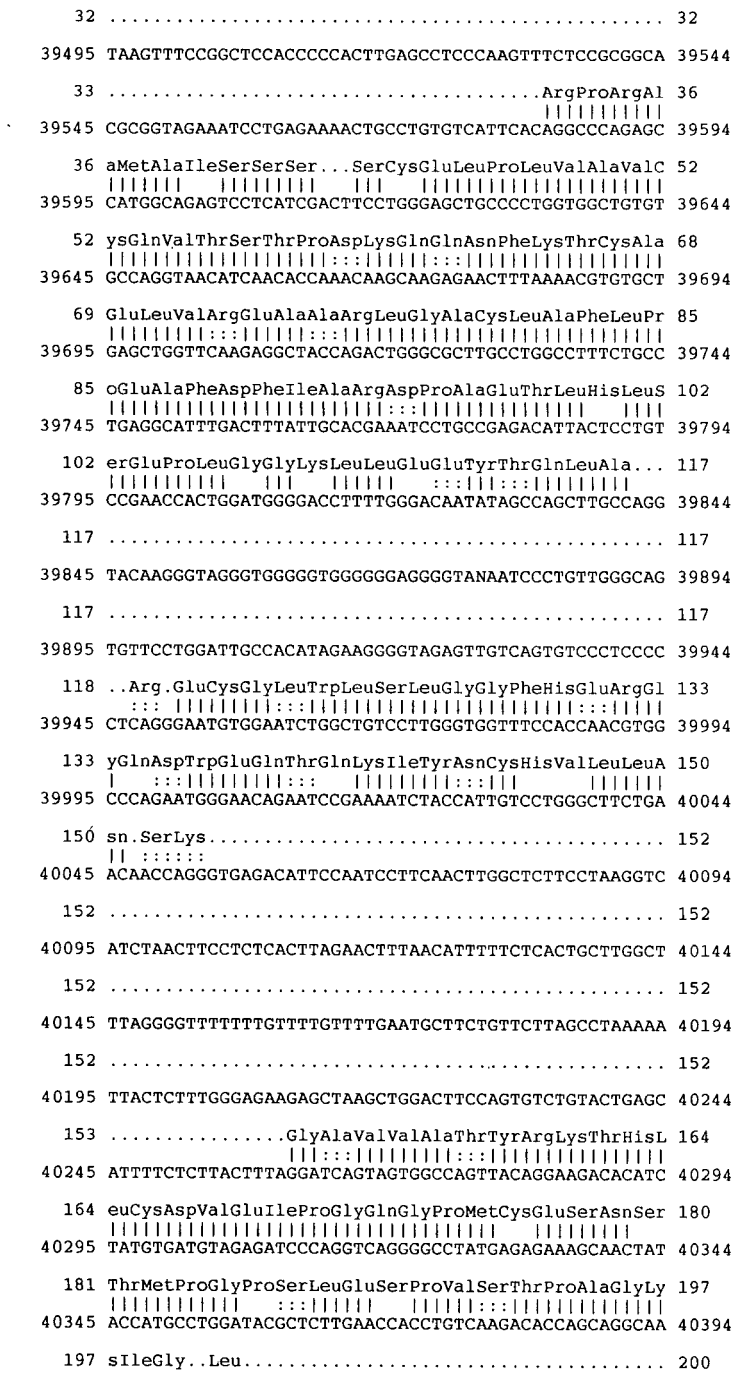
```
 153 .....................................GlyAlaValV 156
                                          |||:::||||
2321 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAGGATCAGTAG 2370

 156 alAlaThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGln 172
     |||||:::||||||||||||||||||||||||||||||||||||||||||
2371 TGGCCAGTTACAGGAAGACACATCTGTGCGATGTAGAGATCCCAGGTCAG 2420

 173 GlyProMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSe 189
     |||||||||   |||||||||   |||   ||||||   :::||||||
2421 GGGCCAATGAGAGAAAGCAACTATACCAAGCCTGGAGGCACTCTTGAGCC 2470

 189 rProValSerThrProAlaGly.Lys..................... 197
      ||||||:::|||||||||||||| :::
2471 ACCTGTCAAGACACCGGCTGGCAAGGTGGGAGTTTGAAAAGGAAATGGGA 2520

 197 ............................................... 197

2521 CTATGTTGAAAAGTGGTAACATCCCCCATTGGTGCAGGCTAATGAAGGCA 2570

 197 ............................................... 197

2571 CTCAGTGTTCTGTTGAATGGGACAGGGAAGATGAGTGGGTTGTTTCTCAT 2620

 198 .....IleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 212
          :::|||||||||:::|||||||||||||||||||||||||||||||
2621 TTCAGGTTGGTCTAGCAATCTGTTATGACATGCGGTTCCCTGAACTTTCT 2670

 213 LeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPh 229
     |||   ||||||||||||||||||||||||||||||||||||||||||||
2671 TTGAAATTGGCTCAAGCTGGGGCAGAAATACTTACTTATCCTTCAGCCTT 2720

 229 eGlySerIleThrGlyProAlaHisTrp..................... 238
     |||||||:::||||||||||||||||||
2721 TGGATCTGTTACAGGTCCAGCCCACTGGGAGGTAAGAGAATGCCCTCTTCA 2770

 238 ............................................... 238

2771 TAACATATAGGCCTCCTTTAACCTTACTTTTAGTAGTGGGAGCAGAGTGG 2820

 238 ............................................... 238

2821 TACAAACCTAAATCCAAGCACTTCAGTGGATAGCCACAGGTTCAAAGCCA 2870

 238 ............................................... 238

2871 GCATGCCTTGACTAAAAAAGAAAACAAAATAGAAGCTGGGCAGTGGGGGT 2920

 238 ............................................... 238

2921 GCACACCTTTAATCCCAGCACTCAGGAGGCAGAAGCAGGTGGATCTCTGA 2970

 238 ............................................... 238

2971 GTTTGAGGCCAGCCTGGTCTACAGAGTGGGTTCCAAGACAGCCAGGGCTA 3020

 238 ............................................... 238

3021 CACAGAGAAACCCTGTCTCAAAAACAAAAAAGCAACCCACACCTTCAAAG 3070

 238 ............................................... 238

3071 AGTGGGGCATGGTGGTGCATGCTCTTCATCTCAGCACTGTGAGGCAAAGG 3120

 238 ............................................... 238

3121 CAGGCATATCTCTGTGGGTCTAGGCCAGCCAGGGTTACACAGGGAGACCT 3170

 238 ............................................... 238

3171 TATCTCAAAAGTACAAAACACAAACACTCCTCCTTTCTACTTGATTTTCT 3220

 238 ............................................... 238

3221 TCCTTTTCATCTGGTGAGGAAATTCATTTCTTAGCTAGGTGTGAGTCATG 3270

 238 ............................................... 238

3271 TTAAAGAGTTAGTTATTAACTCTAGTTGGTTAGTGAGAAGATTGGGCCCA 3320

 238 ............................................... 238

3321 CAGTGCCCGCTTCTTGTCACTTGCCTATTAAACGCTGTATGTTCTGGAGT 3370

 239 .................Glu.ValLeuLeuArgAlaArgAlaIleGluTh 249
                        ::: ||||||||||||||||||||||||||::
3371 CAACGCACGTCTCATCCTCAGGTGCTGCTGCGGGCCCGCGCCATTGAATC 3420

 249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
     :|||||||||||:::|||||||||||||||||||||||||||||   |
3421 TCAGTGCTATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAA 3470

 266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3471 GAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTG 3520

 283 AlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAs 299
     ||||||||||||||||||||||||||||||||||||||||||||||||::
3521 GCCCGCTGCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCA 3570

 299 nTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgP 316
     ::::|||:::||||:::|||:::||||||||||||||||||||||||||
3571 CTTTCTACAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGAC 3620

 316 roAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
     ||||||||||||||:::||||||||||||||||||
3621 CTGACCTCTATGGCAGTCTGGGTCATCCACTCTCT 3655
```

```
seq_name: gb_htg:AC087229

seq_documentation_block:
LOCUS       AC087229              213203 bp    DNA     linear   HTG 05-FEB-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-191a19, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC087229
VERSION     AC087229.15  GI:18497119
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 213203)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Mus musculus Chromosome 1 BAC Clone rp23-191a19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213203)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:18390237.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
           *        1    11373: contig of 11373 bp in length
           *    11374    11473: gap of unknown length
           *    11474    43477: contig of 32004 bp in length
           *    43478    43577: gap of unknown length
           *    43578   126526: contig of 82949 bp in length
           *   126527   126626: gap of unknown length
           *   126627   213203: contig of 86577 bp in length.
FEATURES             Location/Qualifiers
     source          1. .213203
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-191a19"
                     /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT    57858 a  50465 c  49150 g  55425 t    305 others
ORIGIN


alignment_scores:
            Quality: 1083.00            Length:    676
              Ratio:  3.813              Gaps:      9
 Percent Similarity:  42.012   Percent Identity: 37.574

alignment_block:
 US-09-357-675C-21 x AC087229   ..

 Align seg 1/1  to: AC087229  from: 1  to: 213203

      29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGl 45
         |||:::      |||||||||:::|||:::   ||||||:::|||   ||
   48642 TGTGTCATTCACAGGCCCAGAACCATGTCC...TCATCAACTTCCTGGGA 48688

      45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 62
         ||||||||||||||||||||||||||||||||||||||:::||||||:
   48689 GCTGCCCCTGGTGGCTGTGTGCCAGGTAACATCAACACCAAACAAGCAAG 48738

      62 lnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78
         ::|||||||||||||||||||||||||:::||||||||||||||||||
   48739 AGAACTTTAAAACATGTGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGT 48788

      79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr 95
         ||||||||||||||||||||||||||||||||||||||||||:::||
   48789 GCTTGCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCACGAAACCC 48838

      95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112
         ||||||||||||    ||||||||||||||:::|||   ||||||   :
   48839 TGCCGAGACATTACTCCTGTCCGAACCACTGAATGGGGATCTTTTGGGCC 48888

     112 luTyrThrGlnLeuAlaArg..Glu..................... 119
         ::|||:::|||||||||||  :::
   48889 AATATAGCCAGCTTGCCAGGTACAAGGGTAGGGTTTGGGGGGGGGGGGTT 48938

     119 ..................................................... 119

   48939 GTGTGTGTAGAACTCCCTGTTGAGCAGTGTTCCTGGATTGCCACAGAGGG 48988

     120 .....................................CysGlyLe 122
                                              ||||||::
   48989 GGTAGAGTTGTGTGTGTCGTGTCCCCCCCCCCCCCCAGGGAATGTGGAAT 49038

     122 uTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnT 139
         :||||||||||||||||||||||||||||||||||||||||||||||:
   49039 CTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCCAAGACTGGGAGCAGA 49088

     139 hrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAla..V 155
         ::|||||||||||||||||||||||||||||||||||||||||     |
   49089 ATCAGAAAATCTACAATTGTCATGTGCTTTTGAACAGCAAGGGTGAGACG 49138

     155 al.................................................. 155
         ||
   49139 TCATAGCCCTTAGTTTGCCTCTTCCCAGGCTTCTCTACCTTTCCTCAACT 49188

     155 ..................................................... 155

   49189 CAGATCTTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTTTTATGCT 49238

     155 ..................................................... 155

   49239 TCTGTTCTCAGCCTACAAGTTACTCTTTGGGAGAAGAGCTAAGCTGGACT 49288

     155 ..................................................... 155

   49289 TGTAGAGTGTCTGTACTGAGTATTTTCTCTCTCTCTCTCTCTCTCTCT 49338

     156 .................................ValAlaThrTyrAr 160
                                           ||||||:::|||||
   49339 CTCTCTCTCTCTCTCTCTCTCTCTAGGATCAGTAGTGGCCAGTTACAG 49388

     160 gLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysG 177
         ||||||||||||||||||||||||||||||||||||||||||||||   |
   49389 GAAGACACATCTGTGCGATGTAGAGATCCCAGGTCAGGGGCCGATGAGAG 49438

     177 luSerAsnSerThrMetProGlyProSerLeuGluSerProValSerThr 193
         ||||||||   |||   ||||||   :::||||||   ||||||:::|||
   49439 AAAGCAACTATACCAAGCCTGGAGGCACTCTTGAGCCACCTGTCAAGACA 49488

     194 ProAlaGly.Lys..................................... 197
         |||||||||  :::
   49489 CCGGCTGGCAAGGTGGGAGTTTGGAAAGGAAATGGGACTATGTTGAAAAG 49538

     197 ..................................................... 197

   49539 TGGTAACACCCCCCATTGGTGCAGGCTAATGAAGGCACTCAGTGTTCTGT 49588

     198 .....................................IleGlyLe 200
                                               :::||||||
   49589 TGAATGGGACAGGGAAGATGAGTGGGTTGTTTCTCATTTCAGGTTGGTCT 49638

     200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
         ||||:::|||||||||||||||||||||||||||||||   |||||||
   49639 AGCAATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTC 49688

     217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
         |||||||||||||||||||||||||||||||||||||||||||:::|||
   49689 AAGCTGGGGCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTACA 49738

     234 GlyProAlaHisTrp................................. 238
         ||||||||||||||||
   49739 GGTCCAGCCCACTGGGAGGTAAGAGAATGCCTCTTCATAACATATAGGCC 49788

     238 ..................................................... 238

   49789 TCCTTTAACCTTACTTTTAGTAGTGGGAGCAGAGTGGTACAAACCTAAAT 49838

     238 ..................................................... 238

   49839 CCAAGCACTTCAGTGGATAGCCACAGGTTCAAAGCCAGCATGCCTTGACT 49888

     238 ..................................................... 238

   49889 AAAAAAGAAAACAAAATAGAAGCTGGGCAGTGGGGGTGCACACCTTTAAT 49938

     238 ..................................................... 238

   49939 CCCAGCACTCAGGAGGCAGAAGCAGGTGGATCTCTGAGTTTGAGGCCAGC 49988

     238 ..................................................... 238

   49989 CTGGTCTACAGAGTGGGTTCCAGGACAGCCAGGGCTACACAGAGAAACCC 50038

     238 ..................................................... 238

   50039 TGTCTCAAAAACAAAAAAGCAACCCACACCTTCAAAGAGTGGGGCATGGT 50088

     238 ..................................................... 238
```

```
   50089 GGTGCATGCTCTTCATCTCAGCACTGTGAGGCAAAGGCAGGCAGATCTCT 50138

     238 ..................................................... 238

   50139 GTGGGTCTAAGCCAGCCAGGGTTACACAGGGAGACCTTATCTCAAAAGTA 50188

     238 ..................................................... 238

   50189 CAAAACACAAACACTCCTCCTTTCTACTTGATTTTCTTCCTTTTCATCTG 50238

     238 ..................................................... 238

   50239 GTGAGGAAATTCATTTCTTAGCTAGGTGTGAGTCATGTTAAAGAGTTAGT 50288

     238 ..................................................... 238

   50289 TATTAACTCTAGTTGGTTAGTGAGAAGATTGGGCCCACAGTGCCCGCTTC 50338

     238 ..................................................... 238

   50339 TTGTCACTTGCCTATTAAACGCTGTATGTTCTGGAGTCAACGCACGTCTC 50388

     239 ....Glu.ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 253
             ::: |||||||||||||||||||||||||||:::|||||||||||||
   50389 ATCCTCAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAGTGCTATGTA 50438

     254 ValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGl 270
         :::|||||||||||||||||||||||||||||   ||||||||||||||
   50439 ATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGCAAGTTATGG 50488

     270 yHisSerMetValValAspProTrpGlyThrValValAlaArgCysSerG 287
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   50489 CCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGCTCCG 50538

     287 luGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGln 303
         ||||||||||||||||||||||||||||||||||:::::|||:::|||
   50539 AGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTTCTACAACAG 50588

     304 LeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGl 320
         :::|||:::|||||||||||||||||||||||||||||||||||||||||
   50589 ATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGACCTCTATGG 50638

     320 yAsnLeuGlyHisProLeuSer 327
         |:::|||||||||||||||||
   50639 CAGTCTGGGTCATCCACTCTCT 50660

seq_name: gb_htg:AC105589

seq_documentation_block:
LOCUS       AC105589              178418 bp    DNA     linear   HTG 09-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
            ***, 65 unordered pieces.
ACCESSION   AC105589
VERSION     AC105589.1  GI:18092812
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 178418)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178418)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNGI
                Center clone name: CH230-242B2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 132588 bases at least Q40
                Consensus quality: 142233 bases at least Q30
                Consensus quality: 152072 bases at least Q20
                Estimated insert size: 141845; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 65 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8302: contig of 8302 bp in length
            *     8303     8402: gap of unknown length
            *     8403    18472: contig of 10070 bp in length
            *    18473    18572: gap of unknown length
            *    18573    24922: contig of 6350 bp in length
            *    24923    25022: gap of unknown length
            *    25023    30899: contig of 5877 bp in length
            *    30900    30999: gap of unknown length
            *    31000    38320: contig of 7321 bp in length
            *    38321    38420: gap of unknown length
```

```
*    38421    43244: contig of 4824 bp in length
*    43245    43344: gap of unknown length
*    43345    48228: contig of 4884 bp in length
*    48229    48328: gap of unknown length
*    48329    54595: contig of 6267 bp in length
*    54596    54695: gap of unknown length
*    54696    58551: contig of 3856 bp in length
*    58552    58651: gap of unknown length
*    58652    62204: contig of 3553 bp in length
*    62205    62304: gap of unknown length
*    62305    66587: contig of 4283 bp in length
*    66588    66687: gap of unknown length
*    66688    70928: contig of 4241 bp in length
*    70929    71028: gap of unknown length
*    71029    75627: contig of 4599 bp in length
*    75628    75727: gap of unknown length
*    75728    78272: contig of 2545 bp in length
*    78273    78372: gap of unknown length
*    78373    80301: contig of 1929 bp in length
*    80302    80401: gap of unknown length
*    80402    83675: contig of 3274 bp in length
*    83676    83775: gap of unknown length
*    83776    87173: contig of 3398 bp in length
*    87174    87273: gap of unknown length
*    87274    90196: contig of 2923 bp in length
*    90197    90296: gap of unknown length
*    90297    92065: contig of 1769 bp in length
*    92066    92165: gap of unknown length
*    92166    94612: contig of 2447 bp in length
*    94613    94712: gap of unknown length
*    94713    98755: contig of 4043 bp in length
*    98756    98855: gap of unknown length
*    98856   102543: contig of 3688 bp in length
*   102544   102643: gap of unknown length
*   102644   105833: contig of 3190 bp in length
*   105834   105933: gap of unknown length
*   105934   108499: contig of 2566 bp in length
*   108500   108599: gap of unknown length
*   108600   111672: contig of 3073 bp in length
*   111673   111772: gap of unknown length
*   111773   114662: contig of 2890 bp in length
*   114663   114762: gap of unknown length
*   114763   116145: contig of 1383 bp in length
*   116146   116245: gap of unknown length
*   116246   118147: contig of 1902 bp in length
*   118148   118247: gap of unknown length
*   118248   120579: contig of 2332 bp in length
*   120580   120679: gap of unknown length
*   120680   122618: contig of 1939 bp in length
*   122619   122718: gap of unknown length
*   122719   124539: contig of 1821 bp in length
*   124540   124639: gap of unknown length
*   124640   126651: contig of 2012 bp in length
*   126652   126751: gap of unknown length
*   126752   128460: contig of 1709 bp in length
*   128461   128560: gap of unknown length
*   128561   130592: contig of 2032 bp in length
*   130593   130692: gap of unknown length
*   130693   132864: contig of 2172 bp in length
*   132865   132964: gap of unknown length
*   132965   134881: contig of 1917 bp in length
*   134882   134981: gap of unknown length
*   134982   136682: contig of 1701 bp in length
*   136683   136782: gap of unknown length
*   136783   139138: contig of 2356 bp in length
*   139139   139238: gap of unknown length
*   139239   140757: contig of 1519 bp in length
*   140758   140857: gap of unknown length
*   140858   142558: contig of 1701 bp in length
*   142559   142658: gap of unknown length
*   142659   144303: contig of 1645 bp in length
*   144304   144403: gap of unknown length
*   144404   145490: contig of 1087 bp in length
*   145491   145590: gap of unknown length
*   145591   147171: contig of 1581 bp in length
*   147172   147271: gap of unknown length
*   147272   148639: contig of 1368 bp in length
*   148640   148739: gap of unknown length
*   148740   150295: contig of 1556 bp in length
*   150296   150395: gap of unknown length
*   150396   151505: contig of 1110 bp in length
*   151506   151605: gap of unknown length
*   151606   152813: contig of 1208 bp in length
*   152814   152913: gap of unknown length
*   152914   155423: contig of 2510 bp in length
*   155424   155523: gap of unknown length
*   155524   156880: contig of 1357 bp in length
*   156881   156980: gap of unknown length
*   156981   158045: contig of 1065 bp in length
*   158046   158145: gap of unknown length
*   158146   159255: contig of 1110 bp in length
*   159256   159355: gap of unknown length
*   159356   160417: contig of 1062 bp in length
*   160418   160517: gap of unknown length
*   160518   161644: contig of 1127 bp in length
*   161645   161744: gap of unknown length
*   161745   163180: contig of 1436 bp in length
*   163181   163280: gap of unknown length
*   163281   164483: contig of 1203 bp in length
*   164484   164583: gap of unknown length
*   164584   166039: contig of 1456 bp in length
*   166040   166139: gap of unknown length
*   166140   167290: contig of 1151 bp in length

alignment_scores:
           Quality: 1059.00              Length:    835
             Ratio:  3.507                Gaps:     11
 Percent Similarity:  36.168   Percent Identity: 31.976

alignment_block:
 US-09-357-675C-21 x AC105589   ..

 Align seg 1/1  to: AC105589  from: 1  to: 178418

      2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
        ||||||||||||||||||||||||||||::::::|||||||||||||||||
  39095 CTGGGCTTCATCACCAGGCCTCCTCATCAACTCCTGTCCCTTCTGTGTAC 39144

     18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
          |||   |||:::||||||:::||||||||||||:::||||||
  39145 CGGATACCGATTACCTCAAATCTCAGTACTTTGTACTCAGCCCAGGTAAC 39194

     32 ................................................. 32

  39195 ATATTTTATGTTCTTTGTCTAGCACGTAGTTGCTAAGTCCATGTTTATTA 39244

     32 ................................................. 32

  39245 CCCGAATAGTGGACCTGCCATCCACCAAACATTTAATGCGTGTACTCCAG 39294

     32 ................................................. 32

  39295 TAGCTTAGGGTTATTTAATGCTCTGGAAGAATGAAAACAAGTCGCTAGGG 39344

     32 ................................................. 32

  39345 TTGGTGGGACACACCTCTAATCCCAGCACTCGGGAAACAGGCAGGAGGAT 39394

     32 ................................................. 32

  39395 CAACTAGGAGTTTAAGATCATCCCCGCTACATAGAAAGGTCTGCCTGGGC 39444

     32 ................................................. 32

  39445 TACAAAAAGACCCGGTCTCAAAGGGGAGGTAAAATCAGAAAAAAAAAAAG 39494
```


32 32
39495 TAAGTTTCCGGCTCCACCCCCACTTGAGCCTCCCAAGTTTCTCCGCGGCA 39544
33 .....ArgProArgAl 36
39545 CGCGGTAGAAATCCTGAGAAAACTGCCTGTGTCATTCACAGGCCCAGAGC 39594
36 aMetAlaIleSerSerSer...SerCysGluLeuProLeuValAlaValC 52
39595 CATGGCAGAGTCCTCATCGACTTCCTGGGAGCTGCCCCTGGTGGCTGTGT 39644
52 ysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAla 68
39645 GCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACGTGTGCT 39694
69 GluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuPr 85
39695 GAGCTGGTTCAAGAGGCTACCAGACTGGGCGCTTGCCTGGCCTTTCTGCC 39744
85 oGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuS 102
39745 TGAGGCATTTGACTTTATTGCACGAAATCCTGCCGAGACATTACTCCTGT 39794
102 erGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla... 117
39795 CCGAACCACTGGATGGGGACCTTTTGGGACAATATAGCCAGCTTGCCAGG 39844
117 117
39845 TACAAGGGTAGGGTGGGGGTGGGGGGAGGGGTANAATCCCTGTTGGGCAG 39894
117 117
39895 TGTTCCTGGATTGCCACATAGAAGGGGTAGAGTTGTCAGTGTCCCTCCCC 39944
118 ..Arg.GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGl 133
39945 CTCAGGGAATGTGGAATCTGGCTGTCCTTGGGTGGTTTCCACCAACGTGG 39994
133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
39995 CCCAGAATGGGAACAGAATCCGAAAATCTACCATTGTCCTGGGCTTCTGA 40044
150 sn.SerLys..... 152
40045 ACAACCAGGGTGAGACATTCCAATCCTTCAACTTGGCTCTTCCTAAGGTC 40094
152 152
40095 ATCTAACTTCCTCTCACTTAGAACTTTAACATTTTTCTCACTGCTTGGCT 40144
152 152
40145 TTAGGGGTTTTTTTGTTTTGTTTTGAATGCTTCTGTTCTTAGCCTAAAAA 40194
152 152
40195 TTACTCTTTGGGAGAAGAGCTAAGCTGGACTTCCAGTGTCTGTACTGAGC 40244
153 .....GlyAlaValValAlaThrTyrArgLysThrHisL 164
40245 ATTTTCTCTTACTTTAGGATCAGTAGTGGCCAGTTACAGGAAGACACATC 40294
164 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
40295 TATGTGATGTAGAGATCCCAGGTCAGGGGCCTATGAGAGAAAGCAACTAT 40344
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 197
40345 ACCATGCCTGGATACGCTCTTGAACCACCTGTCAAGACACCAGCAGGCAA 40394
197 sIleGly..Leu..... 200
40395 GGTAGGAGTTTGAAAAGAAAATAGGAATATGTTGAAAAAGTTGTAACATC 40444
200 200
40445 TCCCTTTGGTGTAGGATGATGGAGGTAGTATTCAATGTTACGTTGACCAG 40494
201 .....AlaVal 202
40495 GACAGGGAAGATGAGTGGGTTGTTTCTCGTTTCAGGGTGGTCTAGCAATC 40544
203 CysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGl 219
40545 TGTTATGACATGCGGTTCCCTGAACTTTCTCTGAAATGGGCTCAAGCTGG 40594
219 yAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProA 236
40595 GCCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTACCGGCCCAG 40644
236 laHisTrpGluVal..... 240
40645 CCCACTGGGAGGTAGGAGAATGCCTCTTCATAACATATAGGCCTCCTTTA 40694
240 240
40695 ATCTTATTTTTAGTAGTGGGGACAGAGTAGCACAAACCTAAACTCTAAGC 40744
240 240
40745 ACTTCAGGGGATAGCCACAGGTTCAAAGCCAGCATAGTGATGCCTCGAGT 40794
240 240
40795 AAAACAGAAAACAAAATAGAAGCTGGGCAGTGGGGGTGCACACCTTTAAT 40844
240 240
40845 CCCAGCCCTCAGGAGGCAGAAGCAAGTGGATCTGAGTTCAAGACCAGCCT 40894
240 240
40895 GGTCTACAGAGTAGGTTCCAGGACAGCCAGGGCTACACAGAGAAACCGTC 40944
240 240
40945 TCAAAACAAAACAAAACAAAACAAAACAAAAGTAACCCACACCACCCAAA 40994
240 240
40995 ATAGTGGTGCATGGTGGTGTGTGCTTTTTATCTCCGCACTGTAAGGCAGA 41044
240 240
41045 GGCCGGCAGATCTCTGTGGGTCCAGGCCAACCAAGGTTATACAAGGAGAC 41094
240 240
41095 CTTATATAAAAAAAATACAAAACACAAACACACCCCTCCCCTTCTACTTGA 41144
240 240
41145 AGGATTTTCTTCCCTTCCGATCTGGTGAGGAAATTCGTTTCTTAGATAGG 41194
240 240
41195 TGTGAGTCATGTTATAGTTAGTTACTAATTCTAGTTGATCAATAGTCAAT 41244
240 240
41245 GAGTAGATTGGGCCCACAGTGTCCGCTTCTTGTCACTCGCCTATTAAATG 41294
241 .....Leu..LeuArgAla 244

```
  41295 CTGTATGTTCCTGAGTACGCACGTCTCATCCTCAGGTGCTGCTGCGGGCC 41344

    245 ArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyAr 261
        ||||||||||||:::|||||||||||||:::|||||||||||||||||||
  41345 CGCGCCATTGAGTCTCAGTGCTACGTGATAGCAGCAGCACAGTGTGGACG 41394

    261 gHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProT 278
        ||||||||||   ||||||||||||||||||||||||||||||||||||
  41395 CCACCATGAGACAAGAGCAAGTTACGGCCACAGCATGGTGGTTGACCCCT 41444

    278 rpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAla 294
        ||||||||||||||||   ||||||||||||||||||||||||||||||
  41445 GGGGCACAGTGGTGGCCAGCTGTTCCGAGGGACCAGGCCTTTGCCTTGCT 41494

    295 ArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPh 311
        ||||||||||||:::::::|||:::|||:::|||:::|||||||||||||
  41495 CGAATTGATCTCCACTTTCTGCAACAGATGCGCCAACACTTGCCTGTGTT 41544

    311 eGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
        ||||||||||||||||||||||||||||:::|||      |||||||||
  41545 CCAGCATCGCAGACCTGACCTTTATGGCAGTCTC...CTGCCACTCTCC 41590

seq_name: gb_ov:AF284575

seq_documentation_block:
LOCUS       AF284575               1214 bp    mRNA    linear   VRT 22-JUL-2000
DEFINITION  Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION   AF284575
VERSION     AF284575.1  GI:9367117
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1214)
  AUTHORS   Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
  TITLE     Crystal structure of the worm NitFhit Rosetta Stone protein reveals
            a Nit tetramer binding two Fhit dimers
  JOURNAL   Curr. Biol. 10 (15), 907-917 (2000)
  MEDLINE   20414396
REFERENCE   2  (bases 1 to 1214)
  AUTHORS   Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson
            University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1214
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1214
                     /gene="Nit1"
     CDS             180. .1046
                     /gene="Nit1"
                     /note="Nit1; similar to the Nit domains of Caenorhabditis
                     elegans NitFhit and Drosophila melanogaster NitFhit"
                     /codon_start=1
                     /product="Nit protein 1"
                     /protein_id="AAF87104.1"
                     /db_xref="GI:9367118"
                     /translation="MAGAHKPLIAVCQMTSTSDKEKNFATCSRLIREAAGRRACMVFL
                     PEAFDYIGGSIEETLSLAESLHGDTIQRYTQLARECGLWLSLGGFHEKGPNWDTDQRI
                     SNSHVVVDNTGHIVSVYRKAHLFDVDLQNGVSLRESSSTLPGAELIRPITSPAGKIGL
                     GVCYDLRFPEFSLALAQQGAELLTYPSAFTLTTGLAHWEVLLRARAIETQCYVVAAAQ
                     TDRHNEKRTSYGHAMVVDPWGLVIGQCQEGTGICYAEIDIPYMERVRRDMPVWRHRRT
                     DLYGKISFNKPD"
BASE COUNT      281 a    305 c    345 g    282 t      1 others
ORIGIN


alignment_scores:

            Quality:  968.50            Length:    307
              Ratio:  3.905               Gaps:      1
 Percent Similarity:  80.782   Percent Identity: 57.980

alignment_block:
 US-09-357-675C-21 x AF284575   ..

 Align seg 1/1  to: AF284575  from: 1  to: 1214

      16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
         |||||||||   :::   :::   :::|||||||:::::
     144 CTGTGTCCGGATATAGGGGTTAGAAGCCTGAGTGCAATG.......... 182

      32 oArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuV 49
                                        ::::::::::::   ||||||:
     183 ........................GCTGGTGCCCACAAGCCCCTGA 204

      49 alAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLys 65
         ::|||||||||||||:::|||||||||   ||||||:::::|||||||
     205 TTGCCGTGTGCCAGATGACTTCAACCTCTGATAAGGAGAAGAATTTCGCC 254

      66 ThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 82
         ||||||:::::|||:::|||||||||||          ||||||:::::
     255 ACGTGTTCGCGGCTGATCCGGGAGGCTGCGGGGCGTCGCGCTTGCATGGT 304

      82 aPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrL 99
         :|||||||||||||||||||:::|||:::   :::       |||||||
     305 GTTTCTGCCGGAAGCCTTTGACTATATCGGGGGCAGCATTGAGGAGACGC 354

      99 euHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGln 115
         ||   |||:::|||   |||   |||       :::::::::|||||||||
     355 TGAGTCTGGCTGAGTCTCTACATGGGGACACCATTCAGCGTTACACCCAA 404

     116 LeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluAr 132
         ||||||||||||||||||||||||||||||||||||||||||||||||::
     405 CTCGCCAGGGAGTGTGGGCTCTGGCTTTCCCTGGGGGGATTTCATGAGAA 454

     132 gGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuL 149
         :|||   :::|||:::       |||:::|||   |||   ||||||:::
     455 AGGACCCAACTGGGACACGGACCAACGCATTTCCAATTCTCACGTGGTTG 504

     149 euAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCys 165
         ::::::::   |||   :::|||:::::::|||||||||:::||||||
     505 TGGACAACACAGGGCACATAGTATCGGTGTACCGCAAGGCTCACCTGTTT 554

     166 AspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMe 182
         ||||||::::::   :::          :::   ||||||:::||||||::
     555 GACGTAGACTTGCAGAATGGAGTGTCACTCAGAGAGAGCAGTTCCACCCT 604

     182 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199
         :||||||   :::|||       |||:::::::::|||||||||||||||
     605 CCCCGGAGCAGAGCTTATTCGCCCCATCACTTCTCCAGCAGGAAAGATTG 654

     199 lyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeu 215
         |||||:::||||||||||||:::||||||||||||:::||||||||||||
     655 GCCTGGGGGTGTGTTACGACCTCCGCTTCCCAGAATTCTCCTTGGCTCTG 704

     216 AlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIl 232
         ||||||   |||||||||:::||||||||||||||||||||
     705 GCCCAACAAGGAGCAGAACTTCTCACTTACCCTTCTGCCTTCACCCTCAC 754

     232 eThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluT 249
          ||||||   ||||||||||||||||||||||||||||||||||||||||
     755 TACTGGTCTGGCACATTGGGAGGTGTTGCTGAGAGCCCGTGCCATAGAAA 804

     249 hrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLys 265
         ||||||||||||||||||||||||||||||      ||||||:::||||||
     805 CCCAGTGCTACGTAGTTGCAGCGGCACAGACAGACAGACACAATGAGAAG 854

     266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVa 282
         |||:::|||||||||||:::|||||||||||||||||||   |||::
     855 AGGACGTCCTATGGTCACGCTATGGTGGTAGACCCGTGGGGGCTGGTCAT 904
```

```
     282 lAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuA 299
         :::::::|||:::||||||   |||:::|||   |||:::||||||:::
     905 TGGCCAATGCCAGGAAGGAACAGGAATATGTTATGCTGAGATTGACATTC 954

     299 snTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 315
           |||:::::::::::::||||||   :::||||||:::::|||||||||
     955 CCTACATGGAGCGTGTGAGGCGGGACATGCCGGTGTGGAGGCACCGCAGG 1004

     316 ProAspLeuTyrGlyAsnLeu 322
           ||||||||||||:::::
    1005 ACTGATCTGTATGGGAAAATC 1025

seq_name: gb_in:AF069989

seq_documentation_block:
LOCUS       AF069989               1521 bp    mRNA    linear   INV 23-JUL-1998
DEFINITION  Drosophila melanogaster nitrilase and fragile histidine triad
            fusion protein NitFhit (NitFhit) mRNA, complete cds.
ACCESSION   AF069989
VERSION     AF069989.1  GI:3228669
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1521)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and fhit homologs are encoded as fusion proteins in
            drosophila melanogaster and caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
REFERENCE   2  (bases 1 to 1521)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3"
                     /map="61A"
     gene            1. .1521
                     /gene="NitFhit"
     CDS             44. .1426
                     /gene="NitFhit"
                     /codon_start=1
                     /product="nitrilase and fragile histidine triad fusion
                     protein NitFhit"
                     /protein_id="AAC39137.1"
                     /db_xref="GI:3228670"
                     /translation="MSTLVNTTRRSIVIAIHQQLRRMSVQKRKDQSATIAVGQMRSTS
                     DKAANLSQVIELVDRAKSQNACMLFLPECCDFVGESRTQTIELSEGLDGELMAQYREL
                     AKCNKIWISLGGVHERNDQKIFNAHVLLNEKGELAAVYRKLHMFDVTTKEVRLRESDT
                     VTPGYCLERPVSTPVGQIGLQICYDLRFAEPAVLLRKLGANLLTYPSAFTYATGKAHW
                     EILLRARAIETQCFVVAAAQIGWHNQKRQSWGHSMIVSPWGNVLADCSEQELDIGTAE
                     VDLSVLQSLYQTMPCFEHRRNDIYALTAYNLRSKEPTQDRPFATNIVDKRTIFYESEH
                     CFAFTNLRCVVKGHVLVSTKRVTPRLCGLDCAEMADMFTTVCLVQRLLEKIYQTTSAT
                     VTVQDGAQAGQTVPHVHFHIMPRRLGDFGHNDQIYVKLDERAEEKPPRTIEERIEEAQ
                     IYRKFLTDIS"
BASE COUNT      414 a    381 c    391 g    335 t
ORIGIN


alignment_scores:
            Quality:  710.50            Length:    294
              Ratio:  3.259               Gaps:      3
 Percent Similarity:  74.150   Percent Identity: 48.639
```

```
alignment_block:
 US-09-357-675C-21 x AF069989   ..

 Align seg 1/1  to: AF069989  from: 1  to: 1521

      27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
         |||::::::   :::   :::   |||   |||:::::::::::::   ::
      80 GTGATCGCCATTCATCAGCAGCTACGCAGGATGTCCGTCCAAAAGAGGAA 129

      43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
         :   :::          :::||||||   |||:::   ||||||   ||||
     130 GGATCAAAGTGCGACCATTGCTGTGGGTCAGATGCGCTCTACCAGCGACA 179

      60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
         ||       |||::::::          |||||||||   :::|||
     180 AGGCGGCTAATCTTAGCCAAGTGATAGAGCTAGTGGATAGGGCCAAGTCA 229

      77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
            :::||||||:::   ||||||||||||:::   ||||||::::::::
     230 CAGAACGCCTGCATGCTCTTTCTGCCTGAGTGCTGTGACTTTGTGGGCGA 279

      93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
         ::::   ::::::|||::::::|||||||||   |||   |||:::|||:
     280 GAGCCGGACCCAAACTATTGAGCTCTCCGAGGGCTTGGACGGCGAGTTAA 329

     110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
         ::   :::|||   :::||||||:::          :::|||:::||||||
     330 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAAATTTGGATTTCCCTG 379

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
         ||||||   ||||||||||:::::                |||||||||::
     380 GGTGGCGTGCACGAGCGGAACGAT.............CAAAAAATCTT 414

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
         :|||:::|||||||||||||||:::||||||   ::::::|||:::||||
     415 CAACGCTCATGTTTTGCTCAACGAGAAAGGGGAACTAGCAGCAGTATACA 464

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
         |||||   |||:::   ||||||          :::      :::
     465 GAAAGCTGCACATGTTTGATGTTACGACTAAA...GAGGTTCGCCTACGC 511

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
         ||||||:::::::::   ||||||      ||||||   ||||||||||
     512 GAATCAGATACAGTTACGCCGGGATACTGCCTTGAGCGCCCAGTGAGCAC 561

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
         ||||:::|||:::||||||||   :::||||||||||:::||||||   |
     562 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTTGCTG 611

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
         ||   ::::::   |||   :::   ||||||:::::|||||||||||||
     612 AGCCGGCGGTGCTGCTCAGGAAGCTGGGTGCCAATTTGTTAACATACCCA 661

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
         |||||||||        ||||||   ||||||||||||:::|||||||||
     662 TCCGCATTCACATACGCAACCGGTAAGGCGCACTGGGAAATCCTCCTGCG 711

     243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
         ||||||||||||||||||||||||:::|||||||||||||||||||   |
     712 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCGCTGCGGCTCAGATAG 761

     260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
         ||   |||::::::|||||   |||:::||||||||||||:::|||:::
     762 GTTGGCACAACCAGAAGCGACAGAGTTGGGGCCACAGCATGATCGTTAGC 811

     277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
         |||||||||:::|||:::|||   |||||||||           :::
     812 CCCTGGGGAAACGTACTGGCTGACTGCAGCGAGCAGGAGCTTGATATAGG 861

     292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuP 309
            |||::::::||||||:::   |||::::::|||   :::   :::|
     862 CACGGCCGAGGTGGACCTTTCCGTGCTTCAATCTCTGTATCAGACCATGC 911
```

```
    309 roValPheGlnHisArgArgProAspLeuTyr 319
        ||   |||:::|||||||||   |||:::|||
    912 CCTGCTTCGAACATCGTCGAAACGACATCTAC 943

seq_name: gb_htg:AC017755

seq_documentation_block:
LOCUS       AC017755              131853 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017755
VERSION     AC017755.1  GI:6553435
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 131853)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211975 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .131853
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    36037 a  28741 c  28528 g  38547 t
ORIGIN


alignment_scores:
           Quality:  710.50            Length:    294
             Ratio:  3.259               Gaps:      3
 Percent Similarity:  74.150   Percent Identity: 48.639

alignment_block:
 US-09-357-675C-21 x AC017755   ..

 Align seg 1/1  to: AC017755  from: 1  to: 131853

      27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
         |||::::::   :::   :::   |||   |||:::::::::::::   ::
   84569 GTGATCGCCATTCATCAGCAGCTACGCAGGATGTCCGTCCAAAAGAGGAA 84618

      43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
         :  :::          :::||||||   |||:::   ||||||   ||||
   84619 GGATCAAAGTGCGACCATTGCTGTGGGTCAGATGCGCTCTACCAGCGACA 84668

      60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
         ||      |||::::::          |||||||||   :::|||
   84669 AGGCGGCTAATCTTAGCCAAGTGATAGAGCTAGTGGATAGGGCCAAGTCA 84718

      77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
            :::||||||:::   ||||||||||||:::   ||||||:::::::
   84719 CAGAACGCCTGCATGCTCTTTCTGCCTGAGTGCTGTGACTTTGTGGGCGA 84768

      93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
         ::::   :::::::|||:::::::|||||||||   |||   |||:::|||:
   84769 GAGCCGGACCCAAACTATTGAGCTCTCCGAGGGCTTGGACGGCGAGTTAA 84818

     110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
         ::   :::|||   :::|||||||:::          :::|||:::||||||
   84819 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAAATTTGGATTTCCCTG 84868

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143

         ||||||   |||||||||::::::              |||||||||::
   84869 GGTGGCGTGCACGAGCGGAACGAT..............CAAAAAATCTT 84903

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
         :|||:::|||||||||||||||:::||||||   ::::::|||:::||||
   84904 CAACGCTCATGTTTTGCTCAACGAGAAAGGGGAACTAGCAGCAGTATACA 84953

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
         |||||   |||:::   ||||||            :::      :::
   84954 GAAAGCTGCACATGTTTGATGTTACGACTAAA...GAGGTTCGCCTACGC 85000

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
         ||||||:::::::::   ||||||      ||||||   ||||||||||
   85001 GAATCAGATACAGTTACGCCGGGATACTGCCTTGAGCGCCCAGTGAGCAC 85050

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
         ||||:::|||:::|||||||||   :::|||||||||:::||||||   |
   85051 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTTGCTG 85100

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
         ||   ::::::   |||   :::   ||||||::::::||||||||||||
   85101 AGCCGGCGGTGCTGCTCAGGAAGCTGGGTGCCAATTTGTTAACATACCCA 85150

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
         |||||||||         ||||||   ||||||||||||:::|||||||||
   85151 TCCGCATTCACATACGCAACCGGTAAGGCGCACTGGGAAATCCTCCTGCG 85200

     243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
         |||||||||||||||||||||||:::|||||||||||||||||   |
   85201 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCGCTGCGGCTCAGATAG 85250

     260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
         ||   |||::::::||||||   |||:::||||||||||||:::|||:::
   85251 GTTGGCACAACCAGAAGCGACAGAGTTGGGGCCACAGCATGATCGTTAGC 85300

     277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
         |||||||||:::|||:::|||   |||||||||           :::
   85301 CCCTGGGGAAACGTACTGGCTGACTGCAGCGAGCAGGAGCTTGATATAGG 85350

     292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuP 309
         |||::::::||||||:::   |||:::::::|||   :::   :::|
   85351 CACGGCCGAGGTGGACCTTTCCGTGCTTCAATCTCTGTATCAGACCATGC 85400

     309 roValPheGlnHisArgArgProAspLeuTyr 319
         ||   |||:::|||||||||   |||:::|||
   85401 CCTGCTTCGAACATCGTCGAAACGACATCTAC 85432
```

```
OM of: US-09-357-675C-21 to: N_Geneseq_032802:*   out_format : pfs

Date: Apr 29, 2002 10:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091757_3254/app_query.fasta_1.1208
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09357675_@CGN1_1_412 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 338.150000


score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation  ..
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ46101 + 1743.00 2803.93  7.8e-14
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF16257 + 1718.00 2763.69  1.4e-14
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX30398 + 1067.00 1710.69  6.1e-87
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25458 +  968.50 1551.06  4.8e-78
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL12225 +  710.50 1131.04  1.2e-5
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL12224 +  710.50 1122.17  3.7e-5
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20222 +  710.50 1117.33  6.9e-5
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF84214 +  495.00  786.11  1.9e-3
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25456 +  485.50  767.59  2.1e-34
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25457 +  481.00  760.82  5.0e-34
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA81479 -  404.00  595.16  8.4e-25
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF21607 +  404.00  578.65  7.0e-24
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA81490 +  404.00  564.16  4.5e-23
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC77176 +  392.00  613.70  7.8e-26
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL13787 +  387.00  612.29  9.3e-2
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL13786 +  385.00  596.85  6.7e-2
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15102 +  385.00  594.89  8.7e-2
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF11298 +  301.00  477.26  3.1e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH68533 -  292.50  398.05  8.0e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF71819 +  291.00  457.02  4.1e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF72312 +  291.00  457.02  4.1e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH68033 +  283.50  445.87  1.7e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF28550 -  280.50  391.50  1.8e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS82413 +  276.50  425.22  2.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS88625 +  276.50  424.55  2.7e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC50003 +  270.00  420.14  4.7e-15
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC34737 +  270.00  420.12  4.7e-15
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV52205 +  262.00  381.04  7.1e-13
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT30056 -  248.00  368.42  3.6e-12
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS82006 -  229.00  362.53  7.6e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS91910 -  229.00  362.53  7.6e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:AAV74312 -  220.50  323.32  1.2e-09
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH52691 +  218.00  339.82  1.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH54332 +  218.00  324.29  1.0e-0
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC47871 +  217.50  334.55  2.7e-10
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC40019 +  212.50  327.37  6.9e-10
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS41341 +  206.00  321.09  1.5e-0
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH41226 -  206.00  257.95  5.1e-0
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56975 +  205.00  310.46  6.0e-09
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56965 +  204.00  308.84  7.4e-09
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56992 +  204.00  308.84  7.4e
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56993 +  203.00  307.22  9.1e
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56994 +  203.00  307.22  9.1e
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL14462 +  203.00  302.48  1.7
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:AAQ56973 +  202.00  305.61  1.1e

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ46101

seq_documentation_block:
ID   AAZ46101 standard; DNA; 1416 BP.
XX
AC   AAZ46101;
XX
DT   05-MAY-2000  (first entry)
XX
DE   Nucleotide sequence of the coding region of NIT1 gene.
XX
KW   NIT1 gene; nitrilase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW   FRA3B; cancer; genome allele inactivation; ss.
XX
OS   Homo sapiens.
OS   Mus sp.
OS   Drosophila melanogaster.
OS   Caenorhabditis elegans.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1415
FT                   /*tag=  a
FT                   /product= (pos: 18..20, aa: Xaa)
FT                   /product= (pos: 1179..1181, aa: Gln)
FT                   /product= (pos: 1182..1184, aa: Ala)
FT                   /note= "contains 7 internal stop codons; Xaa is an
FT                         unspecified amino acid"
XX
PN   WO200003685-A2.
XX
PD   27-JAN-2000.
XX
PF   20-JUL-1999;   99WO-US16366.
XX
PR   20-JUL-1998;   98US-0093350.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce CM;
XX
DR   WPI; 2000-171195/15.
DR   P-PSDB; AAY68739.
XX
PT   Novel nitrilase homologs used as diagnostic and therapeutic reagents
PT   for the detection and treatment of cancer  -
XX
PS   Claim 6; Fig 6; 25pp; English.
XX
CC   The present sequence represents the coding region of human, murine,
CC   Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC   human and mouse NIT1 genes are members of an uncharacterised
CC   mammalian gene family with homology to bacterial and plant nitrilases.
CC   The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC   for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC   In mouse and humans, FHIT and NIT are encoded by two different genes,
CC   localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC   human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC   chromosomal fragile site FRA3B, is often altered in most common forms
CC   of human cancer. The Nit1 protein overcomes the mutated inactivation
CC   of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC   and analogues of them, and antibodies are used as diagnostic and
CC   therapeutic reagents for the detection and treatment of cancers.
XX
SQ   Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;


alignment_scores:
```

```
          Quality: 1743.00            Length:    326
            Ratio:  5.347               Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-357-675C-21 x AAZ46101   ..

Align seg 1/1  to: AAZ46101  from: 1  to: 1416

      2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    111 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 160

     18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    161 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCA 210

     35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    211 GAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTG 260

     52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    261 TGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGC 310

     68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    311 TGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360

     85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 410

    102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    411 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAG 460

    118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    461 GGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAG 510

    135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    511 ACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGC 560

    152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    561 AAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 610

    168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    611 GATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660

    185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 710

    202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    711 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGC 760

    218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    761 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCC 810

    235 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    811 CAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGC 860

    252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    861 TATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 910

    268 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    911 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

    285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTG 1010

    302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1011 CGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCT 1060

    318 uTyrGlyAsnLeuGlyHisProLeuSer 327
        |||||||||||||||||||||||||||
   1061 CTATGGCAATCTGGGTCACCCACTGTCT 1088

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF16257

seq_documentation_block:
ID   AAF16257 standard; cDNA; 1382 BP.
XX
AC   AAF16257;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB57054.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 1124; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
```

```
SQ   Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;


alignment_scores:
           Quality: 1718.00              Length:    327
             Ratio:  5.270                Gaps:      0
 Percent Similarity:  99.694   Percent Identity: 99.694

alignment_block:
 US-09-357-675C-21 x AAF16257   ..

 Align seq 1/1  to: AAF16257  from: 1  to: 1382

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 126

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 176

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     177 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 226

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     227 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 276

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     277 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 326

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     327 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 426

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     427 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 476

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     477 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 526

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     527 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 576

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         |||||||||||||||||    ||||||||||||||||||||||||||||
     577 AGAGATTCCAGGGCAGGGGCT.ATGTGTGAAAGCAACTCTACCATGCCTG 625

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     626 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 675

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     676 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 725

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     726 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 775

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     776 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 825

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     826 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 875

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     876 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 925

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     926 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 975

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     976 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1025

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1026 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1056

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX30398

seq_documentation_block:
ID   AAX30398 standard; DNA; 1203 BP.
XX
AC   AAX30398;
XX
DT   14-MAY-1999  (first entry)
XX
DE   DNA encoding a human secreted protein.
XX
KW   Secreted protein; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder;
KW   CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW   renal disease; diabetes; inflammation; allergy; ischemic shock;
KW   Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW   prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9907891-A1.
XX
PD   18-FEB-1999.
XX
PF   04-AUG-1998;   98WO-US16235.
XX
PR   19-AUG-1997;   97US-0056732.
PR   05-AUG-1997;   97US-0054798.
PR   05-AUG-1997;   97US-0054803.
PR   05-AUG-1997;   97US-0054804.
PR   05-AUG-1997;   97US-0054806.
PR   05-AUG-1997;   97US-0054807.
PR   05-AUG-1997;   97US-0054808.
PR   05-AUG-1997;   97US-0054809.
PR   05-AUG-1997;   97US-0055309.
PR   05-AUG-1997;   97US-0055310.
PR   05-AUG-1997;   97US-0055312.
PR   05-AUG-1997;   97US-0055386.
PR   05-AUG-1997;   97US-0055311.
PR   18-AUG-1997;   97US-0055970.
PR   18-AUG-1997;   97US-0055986.
PR   19-AUG-1997;   97US-0056365.
PR   19-AUG-1997;   97US-0056366.
PR   19-AUG-1997;   97US-0056557.
PR   19-AUG-1997;   97US-0056370.
PR   19-AUG-1997;   97US-0056371.
PR   19-AUG-1997;   97US-0056563.
PR   19-AUG-1997;   97US-0056731.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Ebner R,  Ferrie AM,  Greene JM,  Janat F,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR,  Young PE,  Yu G;
```

```
XX
DR   WPI; 1999-167452/14.
DR   P-PSDB; AAY10877.
XX
PT   New isolated human genes encoding secreted polypeptides - useful for
PT   diagnosis and treatment of pathalogical diseases
XX
PS   Claim 3; Page 265-266; 331pp; English.
XX
CC   The specification describes secreted proteins and their corresponding
CC   polynucleotides which are useful for preventing, treating or ameliorating
CC   medical conditions, e.g. by protein or gene therapy. Pathological
CC   conditions can also be diagnosed by determining the amount of the
CC   secreted polypeptides in a sample or by determining the presence of
CC   mutations in the polynucleotides. Specific uses are described for each
CC   of the products, based on which tissues they are most highly
CC   expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC   and cognitive disorders, schizophrenia, cardiovascular disorders,
CC   prostate diseases, asthma, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC   thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
XX
SQ   Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;


alignment_scores:
            Quality: 1067.00              Length:    207
              Ratio:  5.205                Gaps:      0
 Percent Similarity:  99.034    Percent Identity: 99.034

alignment_block:
 US-09-357-675C-21 x AAX30398   ..

 Align seg 1/1  to: AAX30398  from: 1  to: 1203

     121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGl 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGA 52

     137 uGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyA 154
         |||||||||||||||||||||||||||||||||||||||   |||||||
      53 GCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAG.AAAGGGG 101

     154 laValValAlaThrTyrArgLysThrHisLeuCysAspValGluIlePro 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCA 151

     171 GlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLe 187
         |||||||||   |||||||||||||||||||||||||||||||||||||||
     152 GGGCAGGGGCT.ATGTGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCT 200

     187 uGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysT 204
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 TGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCT 250

     204 yrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCA 300

     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHi 237
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 GAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCA 350

     237 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValV 254
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAG 400

     254 alAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGly 270
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 TGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGC 450

     271 HisSerMetValValAspProTrpGlyThrValValAlaArgCysSerGl 287
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 CACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGA 500

     287 uGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnL 304
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCGACAGT 550

     304 euArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 TGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGC 600

     321 AsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||
     601 AATCTGGGTCACCCACTGTCT 621

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25458

seq_documentation_block:
ID   AAD25458 standard; cDNA; 1214 BP.
XX
AC   AAD25458;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Xenopus laevis Nit1 cDNA.
XX
KW   Frog; Nit1; cytostatic; neuroprotective; cellular pathway; therapy;
KW   apoptosis; proliferative disorder; degenerative disease; ss.
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
FT   CDS             180..1046
FT                   /*tag=  a
FT                   /product= "Frog Nit1 protein"
XX
PN   WO200187958-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US15664.
XX
PR   16-MAY-2000; 2000US-204713P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce C,  Brenner C,  Pekarski Y;
XX
DR   WPI; 2002-082984/11.
DR   P-PSDB; AAE15792.
XX
PT   Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
PT   useful to find molecules that mimic or antagonize Fhit interaction for
PT   the treatment of proliferative or degenerative diseases
XX
PS   Claim 21; Page 60; 61pp; English.
XX
CC   The invention relates to isolated nucleic acids comprising a fully
CC   defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC   mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion
CC   proteins in invertebrates and as separate polypeptides in vertebrates.
CC   Nit and Fhit interact physically and functions in same cellular pathways.
CC   Molecules which bind Nit2 and mimic or antagonise Fhit interaction are
CC   used to treat diseases in which activity of Nit2 protein is altered in
CC   a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC   treat proliferative disorders, whilst Fhit antagonists promote cell
CC   proliferation and are particularly useful to treat degenerative disease.
CC   The present sequence is frog Nit1 cDNA.
XX
```

```
SQ   Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;


alignment_scores:
           Quality:  968.50              Length:    307
             Ratio:  3.905                Gaps:      1
 Percent Similarity:  80.782   Percent Identity: 57.980

alignment_block:
 US-09-357-675C-21 x AAD25458   ..

 Align seg 1/1  to: AAD25458  from: 1  to: 1214

      16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
         |||||||||   :::   :::   :::||||||::::::
     144 CTGTGTCCGGATATAGGGGTTAGAAGCCTGAGTGCAATG.......... 182

      32 oArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuV 49
                                 ::::::::::::   ||||||:
     183 ..........................GCTGGTGCCCACAAGCCCCTGA 204

      49 alAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLys 65
         ::|||||||||||||:::|||||||||   ||||||:::::|||||||
     205 TTGCCGTGTGCCAGATGACTTCAACCTCTGATAAGGAGAAGAATTTCGCC 254

      66 ThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 82
         ||||||::::::|||:::||||||||||||         ||||||:::::
     255 ACGTGTTCGCGGCTGATCCGGGAGGCTGCGGGGCGTCGCGCTTGCATGGT 304

      82 aPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrL 99
         :|||||||||||||||||||||:::|||:::   :::      |||||||
     305 GTTTCTGCCGGAAGCCTTTGACTATATCGGGGGCAGCATTGAGGAGACGC 354

      99 euHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGln 115
         ||   |||:::|||   |||   |||      :::::::::|||||||||
     355 TGAGTCTGGCTGAGTCTCTACATGGGGACACCATTCAGCGTTACACCCAA 404

     116 LeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluAr 132
         ||||||||||||||||||||||||||||||||||||||||||||||||::
     405 CTCGCCAGGGAGTGTGGGCTCTGGCTTTCCCTGGGGGGATTTCATGAGAA 454

     132 gGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuL 149
         :|||   :::|||:::      |||:::|||   |||   ||||||:::
     455 AGGACCCAACTGGGACACGGACCAACGCATTTCCAATTCTCACGTGGTTG 504

     149 euAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCys 165
         ::::::::  |||   :::|||:::::||||||||:::|||||||
     505 TGGACAACACAGGGCACATAGTATCGGTGTACCGCAAGGCTCACCTGTTT 554

     166 AspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMe 182
         ||||||::::::   :::         :::   ||||||:::||||||::
     555 GACGTAGACTTGCAGAATGGAGTGTCACTCAGAGAGAGCAGTTCCACCCT 604

     182 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199
         :||||||   :::|||      |||:::::::::|||||||||||||||
     605 CCCCGGAGCAGAGCTTATTCGCCCCATCACTTCTCCAGCAGGAAAGATTG 654

     199 lyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeu 215
         |||||:::|||||||||||||:::|||||||||||||:::||||||||||
     655 GCCTGGGGGTGTGTTACGACCTCCGCTTCCCAGAATTCTCCTTGGCTCTG 704

     216 AlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIl 232
         ||||||   |||||||||:::|||||||||||||||||||||
     705 GCCCAACAAGGAGCAGAACTTCTCACTTACCCTTCTGCCTTCACCCTCAC 754

     232 eThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluT 249
          ||||||   ||||||||||||||||||||||||||||||||||||||||
     755 TACTGGTCTGGCACATTGGGAGGTGTTGCTGAGAGCCCGTGCCATAGAAA 804

     249 hrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLys 265
         |||||||||||||||||||||||||||      ||||||:::||||||
     805 CCCAGTGCTACGTAGTTGCAGCGGCACAGACAGACAGACACAATGAGAAG 854

     266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVa 282
         |||:::||||||||||||:::||||||||||||||||||||   |||::
     855 AGGACGTCCTATGGTCACGCTATGGTGGTAGACCCGTGGGGGCTGGTCAT 904

     282 lAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuA 299
         ::::::::|||:::||||||   |||:::|||   |||:::||||||:::
     905 TGGCCAATGCCAGGAAGGAACAGGAATATGTTATGCTGAGATTGACATTC 954

     299 snTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 315
           |||:::::::::::::||||||   :::||||||::::::|||||||||
     955 CCTACATGGAGCGTGTGAGGCGGGACATGCCGGTGTGGAGGCACCGCAGG 1004

     316 ProAspLeuTyrGlyAsnLeu 322
           ||||||||||||::::::
    1005 ACTGATCTGTATGGGAAAATC 1025

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL12225

seq_documentation_block:
ID   ABL12225 standard; cDNA; 1495 BP.
XX
AC   ABL12225;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68122.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31157; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;


alignment_scores:
           Quality:  710.50              Length:    294
             Ratio:  3.259                Gaps:      3
 Percent Similarity:  74.150   Percent Identity: 48.639
```

```
alignment_block:
 US-09-357-675C-21 x ABL12225   ..

 Align seg 1/1  to: ABL12225  from: 1  to: 1495

      27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
      80 GTGATCGCCATTCATCAGCAGCTACGCAGGATGTCCGTCCAAAAGAGGAA 129

      43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
     130 GGATCAAAGTGCGACCATTGCTGTGGGTCAGATGCGCTCTACCAGCGACA 179

      60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
     180 AGGCGGCTAATCTTAGCCAAGTGATAGAGCTAGTGGATAGGGCCAAGTCA 229

      77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
     230 CAGAACGCCTGCATGCTCTTTCTGCCTGAGTGCTGTGACTTTGTGGGCGA 279

      93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
     280 GAGCCGGACCCAAACTATTGAGCTCTCCGAGGGCTTGGACGGCGAGTTAA 329

     110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
     330 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAAATTTGGATTTCCCTG 379

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
     380 GGTGGCGTGCACGAGCGGAACGAT..............CAAAAAATCTT 414

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
     415 CAACGCTCATGTTTTGCTCAACGAGAAAGGGGAACTAGCAGCAGTATACA 464

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
     465 GAAAGCTGCACATGTTTGATGTTACGACTAAA...GAGGTTCGCCTACGC 511

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
     512 GAATCAGATACAGTTACGCCGGGATACTGCCTTGAGCGCCCAGTGAGCAC 561

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
     562 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTTGCTG 611

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
     612 AGCCGGCGGTGCTGCTCAGGAAGCTGGGTGCCAATTTGTTAACATACCCA 661

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
     662 TCCGCATTCACATACGCAACCGGTAAGGCGCACTGGGAAATCCTCCTGCG 711

     243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
     712 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCGCTGCGGCTCAGATAG 761

     260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
     762 GTTGGCACAACCAGAAGCGACAGAGTTGGGGCCACAGCATGATCGTTAGC 811

     277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
     812 CCCTGGGGAAACGTACTGGCTGACTGCAGCGAGCAGGAGCTTGATATAGG 861

     292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuP 309
     862 CACGGCCGAGGTGGACCTTTCCGTGCTTCAATCTCTGTATCAGACCATGC 911

     309 roValPheGlnHisArgArgProAspLeuTyr 319
     912 CCTGCTTCGAACATCGTCGAAACGACATCTAC 943

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL12224

seq_documentation_block:
ID   ABL12224 standard; cDNA; 3548 BP.
XX
AC   ABL12224;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68121.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31154; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;


alignment_scores:
           Quality:  710.50            Length:    294
             Ratio:  3.259               Gaps:      3
 Percent Similarity:  74.150   Percent Identity: 48.639

alignment_block:
 US-09-357-675C-21 x ABL12224   ..

 Align seg 1/1  to: ABL12224  from: 1  to: 3548

      27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
    1133 GTGATCGCCATTCATCAGCAGCTACGCAGGATGTCCGTCCAAAAGAGGAA 1182

      43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
```

```
      :   :::         :::||||||   |||:::   ||||||   ||||
   1183 GGATCAAAGTGCGACCATTGCTGTGGGTCAGATGCGCTCTACCAGCGACA 1232

     60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
        ||      |||::::::         |||||||||   :::|||
   1233 AGGCGGCTAATCTTAGCCAAGTGATAGAGCTAGTGGATAGGGCCAAGTCA 1282

     77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
           :::|||||||:::   |||||||||||||:::   ||||||:::::::::
   1283 CAGAACGCCTGCATGCTCTTTCTGCCTGAGTGCTGTGACTTTGTGGGCGA 1332

     93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
        ::::   ::::::|||::::::|||||||||   |||   |||:::|||:
   1333 GAGCCGGACCCAAACTATTGAGCTCTCCGAGGGCTTGGACGGCGAGTTAA 1382

    110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
        ::   :::|||   :::||||||:::         :::|||:::||||||
   1383 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAAATTTGGATTTCCCTG 1432

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
         ||||||   |||||||||:::::::               |||||||||::
   1433 GGTGGCGTGCACGAGCGGAACGAT.............CAAAAAATCTT 1467

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
         :|||:::|||||||||||||||:::||||||   ::::::|||:::||||
   1468 CAACGCTCATGTTTTGCTCAACGAGAAAGGGGAACTAGCAGCAGTATACA 1517

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
         |||||   |||:::   ||||||           :::      :::
   1518 GAAAGCTGCACATGTTTGATGTTACGACTAAA...GAGGTTCGCCTACGC 1564

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
         ||||||:::::::::   ||||||      ||||||   ||||||||||
   1565 GAATCAGATACAGTTACGCCGGGATACTGCCTTGAGCGCCCAGTGAGCAC 1614

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
         ||||:::|||:::|||||||||   :::|||||||||:::||||||   |
   1615 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTTGCTG 1664

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
         ||   ::::::   |||   :::   ||||||::::::|||||||||||
   1665 AGCCGGCGGTGCTGCTCAGGAAGCTGGGTGCCAATTTGTTAACATACCCA 1714

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
         |||||||||         ||||||   |||||||||||||:::||||||||
   1715 TCCGCATTCACATACGCAACCGGTAAGGCGCACTGGGAAATCCTCCTGCG 1764

     243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
         ||||||||||||||||||||||||:::|||||||||||||||||||   |
   1765 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCGCTGCGGCTCAGATAG 1814

     260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
         ||   |||::::::||||||   |||:::||||||||||||:::|||:::
   1815 GTTGGCACAACCAGAAGCGACAGAGTTGGGGCCACAGCATGATCGTTAGC 1864

     277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
         |||||||||:::|||:::|||   |||||||||            :::
   1865 CCCTGGGGAAACGTACTGGCTGACTGCAGCGAGCAGGAGCTTGATATAGG 1914

     292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuP 309
             |||::::::||||||:::   |||::::::|||   :::   :::|
   1915 CACGGCCGAGGTGGACCTTTCCGTGCTTCAATCTCTGTATCAGACCATGC 1964

     309 roValPheGlnHisArgArgProAspLeuTyr 319
         ||   |||:::||||||||||   |||:::|||
   1965 CCTGCTTCGAACATCGTCGAAACGACATCTAC 1996

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20222

seq_documentation_block:
ID   ABL20222 standard; DNA; 5692 BP.
XX
AC   ABL20222;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;


alignment_scores:
           Quality:  710.50              Length:    294
             Ratio:  3.259                 Gaps:      3
 Percent Similarity:  74.150   Percent Identity: 48.639

alignment_block:
 US-09-357-675C-21 x ABL20222   ..

 Align seg 1/1  to: ABL20222  from: 1  to: 5692

      27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
         |||:::::::   :::   :::   |||   |||:::::::::::::   ::
    2792 GTGATCGCCATTCATCAGCAGCTACGCAGGATGTCCGTCCAAAAGAGGAA 2841

      43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
         :   :::         :::||||||   |||:::   ||||||   ||||
    2842 GGATCAAAGTGCGACCATTGCTGTGGGTCAGATGCGCTCTACCAGCGACA 2891

      60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
         ||      |||::::::         |||||||||   :::|||
    2892 AGGCGGCTAATCTTAGCCAAGTGATAGAGCTAGTGGATAGGGCCAAGTCA 2941

      77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
            :::|||||||:::   |||||||||||||:::   ||||||:::::::::
    2942 CAGAACGCCTGCATGCTCTTTCTGCCTGAGTGCTGTGACTTTGTGGGCGA 2991

      93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
```

```
      ::::   :::::::|||::::::|||||||||   |||   |||:::|||:
   2992 GAGCCGGACCCAAACTATTGAGCTCTCCGAGGGCTTGGACGGCGAGTTAA 3041

    110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
        ::   :::|||   :::|||||||:::         :::|||:::||||||
   3042 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAAATTTGGATTTCCCTG 3091

    127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
        ||||||   |||||||||:::::                |||||||||::
   3092 GGTGGCGTGCACGAGCGGAACGAT.............CAAAAAATCTT 3126

    143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
        :|||:::|||||||||||||||:::||||||   ::::::|||:::||||
   3127 CAACGCTCATGTTTTGCTCAACGAGAAAGGGGAACTAGCAGCAGTATACA 3176

    160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
        |||||   |||:::   ||||||          :::      :::
   3177 GAAAGCTGCACATGTTTGATGTTACGACTAAA...GAGGTTCGCCTACGC 3223

    177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
        ||||||:::::::::   ||||||      ||||||   ||||||||||
   3224 GAATCAGATACAGTTACGCCGGGATACTGCCTTGAGCGCCCAGTGAGCAC 3273

    193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
        ||||:::|||:::|||||||||   :::|||||||||:::||||||   |
   3274 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTTGCTG 3323

    210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
        ||   ::::::   |||   :::   ||||||::::::|||||||||||
   3324 AGCCGGCGGTGCTGCTCAGGAAGCTGGGTGCCAATTTGTTAACATACCCA 3373

    227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
        |||||||||         ||||||   ||||||||||||:::||||||||
   3374 TCCGCATTCACATACGCAACCGGTAAGGCGCACTGGGAAATCCTCCTGCG 3423

    243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
        ||||||||||||||||||||||||:::||||||||||||||||||   |
   3424 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCGCTGCGGCTCAGATAG 3473

    260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
        ||   |||::::::||||||   |||:::|||||||||||||:::|||:::
   3474 GTTGGCACAACCAGAAGCGACAGAGTTGGGGCCACAGCATGATCGTTAGC 3523

    277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
        |||||||||:::|||:::|||   |||||||||            :::
   3524 CCCTGGGGAAACGTACTGGCTGACTGCAGCGAGCAGGAGCTTGATATAGG 3573

    292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuP 309
            |||::::::||||||:::   |||:::::|||   :::   :::|
   3574 CACGGCCGAGGTGGACCTTTCCGTGCTTCAATCTCTGTATCAGACCATGC 3623

    309 roValPheGlnHisArgArgProAspLeuTyr 319
        ||   |||:::|||||||||   |||:::|||
   3624 CCTGCTTCGAACATCGTCGAAACGACATCTAC 3655

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF84214

seq_documentation_block:
ID   AAF84214 standard; cDNA; 1001 BP.
XX
AC   AAF84214;
XX
DT   12-JUN-2001  (first entry)
XX
DE   Human hNit3-ase coding sequence.
XX
KW   Human; hNit3-ase; Cushings adrenal gland tumour;
KW   nitrile hydrolytic enzyme; nitrilase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             99..929
FT                   /*tag=  a
FT                   /product= "Human hNit3-ase"
XX
PN   CN1277998-A.
XX
PD   27-DEC-2000.
XX
PF   30-MAY-2000; 2000CN-0116221.
XX
PR   30-MAY-2000; 2000CN-0116221.
XX
PA   (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.
XX
PI   Xu X,  Qian B,  Zhang X;
XX
DR   WPI; 2001-245678/26.
DR   P-PSDB; AAB80984.
XX
PT   Human nitrile hydrolytic enzyme protein and its coding sequence -
XX
PS   Claim 1; Page 16-17; 20pp; Chinese.
XX
CC   The present invention relates to human hNit3-ase (nitrilase) protein,
CC   which is expressed in Cushings adrenal gland tumour, and its coding
CC   sequence (AAF84214 and AAB80984). The present invention also relates to a
CC   preparation method of said protein and nucleic acid sequence, and a
CC   method of detecting human hNit3-ase nucleic acid sequence and polypeptide
CC   in sample.
XX
SQ   Sequence 1001 BP; 264 A; 235 C; 245 G; 257 T; 0 other;


alignment_scores:
          Quality:  495.00            Length:    321
            Ratio:  2.403               Gaps:      8
 Percent Similarity:  64.174   Percent Identity: 34.891

alignment_block:
 US-09-357-675C-21 x AAF84214   ..

 Align seg 1/1  to: AAF84214  from: 1  to: 1001

      26 SerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSe 42
         ::::::::::::::::::      ||||||:::      ||||||
      11 GCCATTATGGCCGGGGAGGCGACGCCGCGGTCCGCGCTGATCTCCAGGCG 60

      42 rSerCysGluLeu.ProLeuValAlaValCysGlnValThrSer..... 56
          |||   :::       :::|||   ||||||:::|||   :::
      61 CTCAGTCCGCGCCGCAGGTGGTGCTTGTCTGCAGAGTCATGACCTCTTTC 110

      57 .........................ThrProAspLysGlnGlnAsnPh 64
                                     :::      |||::::::|||
     111 CGCTTGGCCCTCATCCAGCTTCAGATTTCTTCCATCAAATCAGATAACGT 160

      64 eLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysL 81
             ::::::::::::::::|||||||||||      ||||||   :
     161 CACTCGCGCTTGTAGCTTCATCCGGGAGGCAGCAACGCAAGGAGCCAAAA 210

      81 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGlu 97
         :::::   |||||||||:::|||
     211 TAGTTTCTTTGCCGGAATGCTTT.......................... 233

      98 ThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyr.. 113
                 ::::::|||   |||:::|||   :::   ||||||
     234 ............AATTCTCCATATGGAGCGAAATATTTTCCTGAATATGC 271

     114 ................................ThrGlnLeuAlaArgG 119
                                         :::::::::|||:::|
     272 AGAGAAAATTCCTGGTGAATCCACACAGAAGCTTTCTGAAGTAGCAAAGG 321

     119 luCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAsp 135
         |||||:::::::::|||      |||::::::   |||
     322 AATGCAGCATATATCTCATTGGAGGCTCTATCCCTGAA............ 359
```

```
     136 TrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLy 152
           |||:::::::   |||:::||||||      ::::::::::::
     360 ...GAGGATGCTGGGAAATTATATAACACCTGTGCTGTGTTTGGGCCTGA 406

     152 sGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGluI 169
          |||::::::::::|||   |||||||||   ||||||   |||:::::::
     407 TGGAACTTTACTAGCAAAGTATAGAAAGATCCATCTGTTTGACATTGATG 456

     169 leProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyPro 185
         ::||||||:::      :::   ||||||::::::      ||||||
     457 TTCCTGGAAAAATTACATTTCAAGAATCTAAAACATTGAGTCCGGGTGAT 506

     186 SerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaVa 202
         |||:::   |||      :::||||||      ::::::||||||:::::
     507 AGTTTC...TCCACATTTGATACTCCTTACTGCAGAGTGGGTCTGGGCAT 553

     202 lCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaG 219
         :|||||||||||||||||   ||||||:::         ||||||   |
     554 CTGCTACGACATGCGGTTTGCAGAGCTTGCACAAATCTACGCACAGAGAG 603

     219 lyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyPro 235
         ||:::::::::|||:::||||||:::||||||:::      |||||||||
     604 GCTGCCAGCTGTTGGTATATCCAGGAGCTTTTAATCTGACCACTGGACCA 653

     236 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTy 252
         ||||||||||||:::|||   |||:::||||||:::::::::|||   ||
     654 GCCCATTGGGAGTTACTTCAGCGAAGCCGGGCTGTTGATAATCAGGTGTA 703

     252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
         ||||::::::|||:::   :::        ::::::   :::|||||||
     704 TGTGGCCACAGCCTCTCCTGCC.........CGGGATGACAAAGCCTCCT 744

     269 yr.........GlyHisSerMetValValAspProTrpGlyThrValVal 282
         ||         |||||||||   ||||||:::|||||||||   |||:::
     745 ATGTTGCCTGGGGACACAGCACCGTGGTGAACCCTTGGGGGGAGGTTCTA 794

     283 AlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAs 299
         |||::::::::::         ::::::      :::   |||||||||::
     795 GCCAAAGCTGGCACAGAAGAAGCAATCGTGTATTCAGACATAGACCTGAA 844

     299 nTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgP 316
         :   |||   ::::::|||::::::::::|||||||||:::::::::|||
     845 GAAGCTGGCTGAAATACGCCAGCAAATCCCCGTTTTTAGACAGAAGCGAT 894

     316 roAspLeuTyr 319
           |||||||||
     895 CAGACCTCTAT 905

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25456

seq_documentation_block:
ID   AAD25456 standard; cDNA; 1359 BP.
XX
AC   AAD25456;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human Nit2 cDNA.
XX
KW   Human; Nit2; cytostatic; neuroprotective; cellular pathway; therapy;
KW   apoptosis; proliferative disorder; degenerative disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..852
FT                   /*tag=  a
FT                   /product= "Human Nit2 protein"
XX
PN   WO200187958-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US15664.
XX
PR   16-MAY-2000; 2000US-204713P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce C,  Brenner C,  Pekarski Y;
XX
DR   WPI; 2002-082984/11.
DR   P-PSDB; AAE15790.
XX
PT   Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
PT   useful to find molecules that mimic or antagonize Fhit interaction for
PT   the treatment of proliferative or degenerative diseases  -
XX
PS   Claim 3; Page 59; 61pp; English.
XX
CC   The invention relates to isolated nucleic acids comprising a fully
CC   defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC   mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion
CC   proteins in invertebrates and as separate polypeptides in vertebrates.
CC   Nit and Fhit interact physically and functions in same cellular pathways.
CC   Molecules which bind Nit2 and mimic or antagonise Fhit interaction are
CC   used to treat diseases in which activity of Nit2 protein is altered in
CC   a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC   treat proliferative disorders, whilst Fhit antagonists promote cell
CC   proliferation and are particularly useful to treat degenerative disease.
CC   The present sequence is human Nit2 cDNA.
XX
SQ   Sequence 1359 BP; 388 A; 284 C; 312 G; 374 T; 1 other;


alignment_scores:
            Quality:  485.50              Length:    286
              Ratio:  2.542                 Gaps:      6
 Percent Similarity:  66.783   Percent Identity: 36.713

alignment_block:
 US-09-357-675C-21 x AAD25456   ..

 Align seg 1/1  to: AAD25456  from: 1  to: 1359

      49 ValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLy 65
         :::|||:::   |||:::      :::      |||:::::::|||
      37 TTGGCCCTCATCCAGCTTCAGATTTCTTCCATCAAATCAGATAACGTCAC 86

      65 sThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuA 82
            ::::::::::::::::|||||||||||||      ||||||   ::::
      87 TCGCGCTTGTAGCTTCATCCGGGAGGCAGCAACGCAAGGAGCCAAAATAG 136

      82 laPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 98
         ::   ||||||||||:::|||
     137 TTTCTTTGCCGGAATGCTTT............................... 156

      99 LeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyr..... 113
                ::::::|||   |||:::|||   :::   ||||||
     157 .........AATTCTCCATATGGAGCGAAATATTTTCCTGAATATGCAGA 197

     114 ...............................ThrGlnLeuAlaArgGluC 120
                                        :::::::::|||:::||||
     198 GAAAATTCCTGGTGAATCCACACAGAAGCTTTCTGAAGTAGCAAAGGAAT 247

     120 ysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrp 136
         ||:::::::::|||      |||::::::   |||
     248 GCAGCATATATCTCATTGGAGGCTCTATCCCTGAA............... 282

     137 GluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGl 153
         |||::::::   |||:::||||||      ::::::::::::      ||
     283 GAGGATGCTGGGAAATTATATAACACCTGTGCTGTGTTTGGGCCTGATGG 332

     153 yAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGluIleP 170
         |:::::::::|||   |||||||||   ||||||   |||::::::::::|
```

```
     333 AACTTTACTAGCAAAGTATAGAAAGATCCATCTGTTTGACATTGATGTTC 382

     170 roGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyProSer 186
         |||||:::      :::   ||||||::::::      ||||||   |||
     383 CTGGAAAAATTACATTTCAAGAATCTAAAACATTGAGTCCGGGTGATAGT 432

     187 LeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCy 203
         :::   |||      :::||||||      ::::::||||||::::::||
     433 TTC...TCCACATTTGATACTCCTTACTGCAGAGTGGGTCTGGGCATCTG 479

     203 sTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyA 220
         ||||||||||||||||   ||||||:::         ||||||   |||:
     480 CTACGACATGCGGTTTGCAGAGCTTGCACAAATCTACGCACAGAGAGGCT 529

     220 laGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAla 236
         :::::::::|||:::||||||:::||||||:::      ||||||||||||
     530 GCCAGCTGTTGGTATATCCAGGAGCTTTTAATCTGACCACTGGACCAGCC 579

     237 HisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa 253
         |||||||||:::|||   |||:::||||||:::::::::|||   |||||
     580 CATTGGGAGTTACTTCAGCGAAGCCGGGCTGTTGATAATCAGGTGTATGT 629

     253 lValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyr. 269
         |:::::|||:::   :::         ::::::   :::|||||||||
     630 GGCCACAGCCTCTCCTGCC.........CGGGATGACAAAGCCTCCTATG 670

     270 ........GlyHisSerMetValValAspProTrpGlyThrValValAla 283
                 |||||||||   ||||||:::|||||||||   |||:::|||
     671 TTGCCTGGGGACACAGCACCGTGGTGAACCCTTGGGGGGAGGTTCTAGCC 720

     284 ArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTy 300
         :::::::::         ::::::      :::   |||||||||:::
     721 AAAGCTGGCACAGAAGAAGCAATCGTGTATTCAGACATAGACCTGAAGAA 770

     300 rLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProA 317
          |||   :::::::|||:::::::::|||||||||:::::::::|||   |
     771 GCTGGCTGAAATACGCCAGCAAATCCCCGTTTTTAGACAGAAGCGATCAG 820

     317 spLeuTyr 319
         ||||||||
     821 ACCTCTAT 828

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAD25457

seq_documentation_block:
ID   AAD25457 standard; cDNA; 1292 BP.
XX
AC   AAD25457;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Mouse Nit2 cDNA.
XX
KW   Mouse; Nit2; cytostatic; neuroprotective; cellular pathway; therapy;
KW   apoptosis; proliferative disorder; degenerative disease; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             46..876
FT                   /*tag=  a
FT                   /product= "Mouse Nit2 protein"
XX
PN   WO200187958-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US15664.
XX
PR   16-MAY-2000; 2000US-204713P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce C,  Brenner C,  Pekarski Y;
XX
DR   WPI; 2002-082984/11.
DR   P-PSDB; AAE15791.
XX
PT   Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
PT   useful to find molecules that mimic or antagonize Fhit interaction for
PT   the treatment of proliferative or degenerative diseases  -
XX
PS   Claim 9; Page 59-60; 61pp; English.
XX
CC   The invention relates to isolated nucleic acids comprising a fully
CC   defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC   mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion
CC   proteins in invertebrates and as separate polypeptides in vertebrates.
CC   Nit and Fhit interact physically and functions in same cellular pathways.
CC   Molecules which bind Nit2 and mimic or antagonise Fhit interaction are
CC   used to treat diseases in which activity of Nit2 protein is altered in
CC   a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC   treat proliferative disorders, whilst Fhit antagonists promote cell
CC   proliferation and are particularly useful to treat degenerative disease.
CC   The present sequence is mouse Nit2 cDNA.
XX
SQ   Sequence 1292 BP; 354 A; 292 C; 289 G; 357 T; 0 other;


alignment_scores:
            Quality:  481.00              Length:    302
              Ratio:  2.324                 Gaps:      9
 Percent Similarity:  68.543   Percent Identity: 37.748

alignment_block:
 US-09-357-675C-21 x AAD25457   ..

 Align seg 1/1  to: AAD25457  from: 1  to: 1292

      24 GlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSe 40
         |||   |||:::      |||      |||   |||||||||
       7 CAGGAGTCCCTGTGGCCTGCGGTGCACCGTGGAAGAGCCATG........ 48

      40 rSerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerT 57
             |||:::   :::|||   |||:::   :::   ||||||:::|||
      49 ....TCTACTTTCCGCCTGGCCCTCATACAGCTT...CAAGTTTCTTCCA 91

      57 hrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGlu 73
                |||::::::|||:::     :::::::::||||||||||||
      92 TT......AAATCAGATAACCTTACCCGGGCTTGTAGCCTAGTGCGGGAG 135

      74 AlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPh 90
         ||||||:::   ||||||   ::::::   ||||||||||:::|||:::
     136 GCAGCAAAGCAAGGTGCCAACATAGTTTCTCTGCCTGAGTGCTTCAAT.. 183

      90 eIleAlaArgAspProAlaGluThrLeuHisLeu.........SerGluP 104
                  :::|||      |||   ::::::         :::|||
     184 ..........TCTCCATATGGAACAACCTACTTTCCTGACTATGCAGAGA 223

     104 roLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCys 120
           :::   |||:::      ::::::   :::::::::|||:::|||
     224 AGATTCCTGGAGAGTCCACACAAAAGCTTTCTGAAGTAGCAAAGGAGAGC 273

     121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGl 137
         ::::::::::|||      |||::::::   |||               ||
     274 AGCATATATCTCATTGGAGGCTCCATCCCTGAA..............GA 308

     137 uGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyA 154
         |::::::   |||:::||||||         :::::::::      |||:
     309 GGATGCTGGGAAACTGTATAATACCTGCTCTGTGTTTGGGCCTGATGGAA 358

     154 laValValAlaThrTyrArgLysThrHisLeuCysAspValGluIlePro 170
         ::::::::::::   :::||||||   ||||||   |||:::::::::|||
     359 GTTTACTGGTAAAGCACAGGAAGATCCATCTGTTTGACATTGATGTTCCT 408

     171 GlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLe 187
```

```
     409 GGGAAAATTACGTTTCAAGAATCTAAAACATTGAGCCCTGGTGATAGTTT 458

     187 uGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysT 204
     459 C...TCCACATTTGATACGCCTTACTGCAAAGTGGGCCTGGGCATCTGCT 505

     204 yrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
     506 ATGATATGCGCTTCGCGGAGCTTGCACAAATCTATGCACAAAGAGGCTGC 555

     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHi 237
     556 CAGCTCTTGGTGTATCCTGGAGCTTTCAATCTGACCACAGGACCAGCCCA 605

     237 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValV 254
     606 CTGGGAGCTGCTTCAGCGAGCCCGGGCTGTTGATAATCAGGTGTATGTGG 655

     254 alAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyr... 269
     656 CTACAGCCTCTCCTGCT.........CGGGATGACAAAGCCTCGTATGTG 696

     270 ......GlyHisSerMetValValAspProTrpGlyThrValValAlaAr 284
     697 GCCTGGGGACACAGCACTGTTGTGGATCCTTGGGGGCAGGTCCTAACCAA 746

     284 gCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
     747 AGCTGGCACGGAGGAAACAATCCTGTACTCAGACATAGACCTGAAGAAGC 796

     301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
     797 TGGCTGAAATTCGGCAGCAAATCCCCATTTTAAAACAGAAACGAGCAGAC 846

     318 LeuTyr 319
     847 CTCTAT 852

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA81479

seq_documentation_block:
ID   AAA81479 standard; DNA; 69936 BP.
XX
AC   AAA81479;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 547-567; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

alignment_scores:
          Quality:  404.00              Length:    307
            Ratio:  2.138                 Gaps:      8
 Percent Similarity:  61.564   Percent Identity: 35.179

alignment_block:
 US-09-357-675C-21 x AAA81479/rev   ..

 Align seg 1/1  to reverse of: AAA81479  from: 1  to: 69936

      16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
   27227 TTGTGTAAA......CGGTTTTCGCAAAGTACCGTCATGGTAGCCTATCG 27184

      32 oArg......ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
   27183 GCGGAATATGCCGCAAGGTCGGCAGGAAAAAGGAGAAGAAATGGACAAAA 27134

      47 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 63
   27133 TCAGAGTTGCCGCCGTGCAGATGGTGTCGGGCGTGTCGCCGGAAACCAAC 27084

      64 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCy 80
   27083 GTCGCCGCCATGAAACGCCTGGTCGCACGGGCGGCGGAGCAGGGTGCGGA 27034

      80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
   27033 TTGGGTGCTGCTGCCCGAATATTGGGTGCTGATGGGCGCAAACGATACCG 26984

      97 luThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuLeuGluGlu 112
   26983 ACAAACTCGCGCTTGCCGAGCCTTTGGGCGGCGGACGCTTTCAGACGGCA 26934

     113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
   26933 TTGAGCGAAACGGCGAAAGAATGCGGCGTGGTGCTG...TTCGGCGGGAC 26887
```

```
  129 .............PheHisGluArgGlyGlnAspTrpGluGlnThrGlnL 141
                      |||   |||                        |
26886 TGTGCCGCTGCAAAGCTGCGAGGCGGGT.....................A 26858

  141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
      ||:::   |||      ::::::   :::      |||:::   ::::::
26857 AAGTGATGAATACGCTGTTGGTGTACGGACGGGACGGCGTAAGGACGGGG 26808

  158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
         |||:::|||   ||||||            :::   |||   |||
26807 CTGTACCACAAAATGCACCTC......TTCGGTTTTTCCGGTTTGGGCGA 26764

  174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
              :::   :::|||:::   :::   :::             :
26763 ACGCTATGCCGAAGCCGATACCATCCGCGCGGGCGGGGATGTGCCGCACT 26714

  191 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 207
      ::|||:::   :::     ::::::   ::::::|||||||||:::|||
26713 TGTCGGCAGAAGGCGTGCCGGTGGCGGCGGGCATTTGTTACGATGTCCGC 26664

  208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
      |||||||||      :::         |||         :::::::|||
26663 TTTCCCGAA......TTTTTCCGACGCCAGTTGCCGTTTGACGTATTGAT 26620

  224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValL 241
          |||:::||||||         ||||||   ||||||||||||:::|
26619 GCTGCCCGCTGCGTTTACGCACACGACGGGCAAGGCGCATTGGGAGCTGC 26570

  241 euLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAla 257
      ||||||||||||||||:::|||:::|||||||||||||||||||||||||
26569 TGCTGCGCGCGCGTGCCGTCGAAAACCAATGTTACGTCGTGGCGGCGGCA 26520

  258 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVa 274
      |||   |||   |||::::::   |||   :::::|||||||||||||::
26519 CAGGGCGGTTTGCACGAAAACGGACGGCGCACGTTCGGACACAGCATGAT 26470

  274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
      :|||||||||||||||   |||:::           ||||||   |||:
26469 TGTCGATCCGTGGGGCGACGTGTTGGACGTATTGCCCGAGGGCGAAGGCG 26420

  291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
      ::      |||   ||||||   |||   |||::::::::::|||::::::
26419 TTGTTACGGCAGACATCGATGCCAACCGCCTGAACAGCGTCCGCAACCGC 26370

  308 LeuProValPheGlnHisArg 314
      ||||||::::::::::::|||
26369 CTGCCCGCCTTGAAATACCGG 26349
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAF21607

seq_documentation_block:
ID   AAF21607 standard; DNA; 349980 BP.
XX
AC   AAF21607;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;
```

```
alignment_scores:
           Quality:  404.00              Length:    307
             Ratio:  2.138                Gaps:      8
 Percent Similarity:  61.564   Percent Identity: 35.179

alignment_block:
 US-09-357-675C-21 x AAF21607   ..

 Align seg 1/1  to: AAF21607  from: 1  to: 349980

     16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
        ||||||         ||:::   |||   :::|||:::   |||
 153960 TTGTGTAAA......CGGTTTTCGCAAAGTACCGTCATGGTAGCCTATCG 154003

     32 oArg......ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
          |||      |||::::::         ::::::::::::   :::
 154004 GCGGAATATGCCGCAAGGTCGGCAGGAAAAAGGAGAAGAAATGGACAAAA 154053

     47 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 63
           ||||||:::   |||::::::|||      :::   :::   |||
 154054 TCAGAGTTGCCGCCGTGCAGATGGTGTCGGGCGTGTCGCCGGAAACCAAC 154103

     64 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCy 80
           :::      :::||||||   :::||||||:::   ||||||
 154104 GTCGCCGCCATGAAACGCCTGGTCGCACGGGCGGCGGAGCAGGGTGCGGA 154153

     80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
         ::::::||||||||||   :::   :::::::::   :::   ::::
 154154 TTGGGTGCTGCTGCCCGAATATTGGGTGCTGATGGGCGCAAACGATACCG 154203

     97 luThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuLeuGluGlu 112
```

```
       ::   |||   |||:::|||||||||   ||||||::::::
154204 ACAAACTCGCGCTTGCCGAGCCTTTGGGCGGCGGACGCTTTCAGACGGCA 154253

   113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
       ::::::   |||:::||||||||||:::   |||   :::|||||||
154254 TTGAGCGAAACGGCGAAAGAATGCGGCGTGGTGCTG...TTCGGCGGGAC 154300

   129 .............PheHisGluArgGlyGlnAspTrpGluGlnThrGlnL 141
                       |||   |||                       |
154301 TGTGCCGCTGCAAAGCTGCGAGGCGGGT.....................A 154329

   141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
       ||:::   |||      ::::::   :::      |||:::   ::::::
154330 AAGTGATGAATACGCTGTTGGTGTACGGACGGGACGGCGTAAGGACGGGG 154379

   158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
       |||:::|||   ||||||          :::   |||   |||
154380 CTGTACCACAAAATGCACCTC......TTCGGTTTTTCCGGTTTGGGCGA 154423

   174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
               :::   :::|||:::   :::   :::            :
154424 ACGCTATGCCGAAGCCGATACCATCCGCGCGGGCGGGGATGTGCCGCACT 154473

   191 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 207
       ::|||:::   :::     ::::::   ::::::|||||||||:::|||
154474 TGTCGGCAGAAGGCGTGCCGGTGGCGGCGGGCATTTGTTACGATGTCCGC 154523

   208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
       |||||||||      :::   ·      |||         :::::|||
154524 TTTCCCGAA......TTTTTCCGACGCCAGTTGCCGTTTGACGTATTGAT 154567

   224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValL 241
           |||:::||||||         ||||||   ||||||||||||:::|
154568 GCTGCCCGCTGCGTTTACGCACACGACGGGCAAGGCGCATTGGGAGCTGC 154617

   241 euLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAla 257
       ||||||||||||||||:::|||:::||||||||||||||||||||||||
154618 TGCTGCGCGCGCGTGCCGTCGAAAACCAATGTTACGTCGTGGCGGCGGCA 154667

   258 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVa 274
       |||   |||   |||::::::   |||   ::::::||||||||||||::
154668 CAGGGCGGTTTGCACGAAAACGGACGGCGCACGTTCGGACACAGCATGAT 154717

   274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
       :|||||||||||||||   |||:::            ||||||   |||:
154718 TGTCGATCCGTGGGGCGACGTGTTGGACGTATTGCCCGAGGGCGAAGGCG 154767

   291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
       ::      |||   ||||||   |||   |||:::::::::|||::::::
154768 TTGTTACGGCAGACATCGATGCCAACCGCCTGAACAGCGTCCGCAACCGC 154817

   308 LeuProValPheGlnHisArg 314
       ||||||:::::::::::::|||
154818 CTGCCCGCCTTGAAATACCGG 154838

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA81490

seq_documentation_block:
ID   AAA81490 standard; DNA; 1437668 BP.
XX
AC   AAA81490;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 866-1272; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;


alignment_scores:
            Quality:  404.00              Length:    307
              Ratio:  2.138                 Gaps:      8
 Percent Similarity:  61.564    Percent Identity: 35.179

alignment_block:
 US-09-357-675C-21 x AAA81490   ..

 Align seg 1/1  to: AAA81490  from: 1  to: 1437668

     16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
        ||||||         |||:::   |||   :::|||:::   |||
 453960 TTGTGTAAA......CGGTTTTCGCAAAGTACCGTCATGGTAGCCTATCG 454003

     32 oArg......ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
         |||      |||::::::          ::::::::::::   :::
 454004 GCGGAATATGCCGCAAGGTCGGCAGGAAAAAGGAGAAGAAATGGACAAAA 454053

     47 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 63
           ||||||:::   |||:::::::|||      :::   :::   |||
 454054 TCAGAGTTGCCGCCGTGCAGATGGTGTCGGGCGTGTCGCCGGAAACCAAC 454103
```

```
     64 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCy 80
454104 GTCGCCGCCATGAAACGCCTGGTCGCACGGGCGGCGGAGCAGGGTGCGGA 454153

     80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
454154 TTGGGTGCTGCTGCCCGAATATTGGGTGCTGATGGGCGCAAACGATACCG 454203

     97 luThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuGluGlu 112
454204 ACAAACTCGCGCTTGCCGAGCCTTTGGGCGGCGGACGCTTTCAGACGGCA 454253

    113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
454254 TTGAGCGAAACGGCGAAAGAATGCGGCGTGGTGCTG...TTCGGCGGGAC 454300

    129 .............PheHisGluArgGlyGlnAspTrpGluGlnThrGlnL 141
454301 TGTGCCGCTGCAAAGCTGCGAGGCGGGT.....................A 454329

    141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
454330 AAGTGATGAATACGCTGTTGGTGTACGGACGGGACGGCGTAAGGACGGGG 454379

    158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
454380 CTGTACCACAAAATGCACCTC......TTCGGTTTTTCCGGTTTGGGCGA 454423

    174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
454424 ACGCTATGCCGAAGCCGATACCATCCGCGCGGGCGGGGATGTGCCGCACT 454473

    191 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 207
454474 TGTCGGCAGAAGGCGTGCCGGTGGCGGCGGGCATTTGTTACGATGTCCGC 454523

    208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
454524 TTTCCCGAA......TTTTTCCGACGCCAGTTGCCGTTTGACGTATTGAT 454567

    224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValL 241
454568 GCTGCCCGCTGCGTTTACGCACACGACGGGCAAGGCGCATTGGGAGCTGC 454617

    241 euLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAla 257
454618 TGCTGCGCGCGCGTGCCGTCGAAAACCAATGTTACGTCGTGGCGGCGGCA 454667

    258 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVa 274
454668 CAGGGCGGTTTGCACGAAAACGGACGGCGCACGTTCGGACACAGCATGAT 454717

    274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
454718 TGTCGATCCGTGGGGCGACGTGTTGGACGTATTGCCCGAGGGCGAAGGCG 454767

    291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
454768 TTGTTACGGCAGACATCGATGCCAACCGCCTGAACAGCGTCCGCAACCGC 454817

    308 LeuProValPheGlnHisArg 314
454818 CTGCCCGCCTTGAAATACCGG 454838

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC77176

seq_documentation_block:
ID   AAC77176 standard; cDNA; 1725 BP.
XX
AC   AAC77176;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2731 polynucleotide sequence SEQ ID NO:5461.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42967.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 4641-4643; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1725 BP; 466 A; 368 C; 372 G; 517 T; 2 other;

alignment_scores:
           Quality:  392.00              Length:    481
             Ratio:  1.858                 Gaps:     14
 Percent Similarity:  43.867   Percent Identity: 24.324
```

```
alignment_block:
 US-09-357-675C-21 x AAC77176   ..

 Align seg 1/1  to: AAC77176  from: 1  to: 1725

      23 ProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIl 39
      11 CCGCAGGTGGTGCTTGTCTGC..............AGAGTCATG..... 40

      39 eSerSerSerSerCysGluLeuProLeuValAlaValCysGlnValThrS 56
      41 .......ACCTCTTTCCGCTTGGCCCTCATCCAGCTT...CAGATTTCTT 80

      56 erThrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArg 72
      81 CCATC......AAATCAGATAACGTCACTCGCGCTTGTAGCTTCATCCGG 124

      73 GluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAs 89
     125 GAGGCAGCAACGCAAGGAGCCAAAATAGTTTCTTTGCCGGAATGCTTT.. 172

      89 pPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 106
     173 .....................................AATTCTCCATATG 185

     106 lyGlyLysLeuLeuGluGluTyr..................... 113
     186 GAGCGAAATATTTTCCTGAATATGCAGAGAAAATTCCTGGTGAATCCACA 235

     114 .........ThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
     236 CAGAAGCTTTCTGAAGTAGCAAAGGAATGCAGCATATATCTCATTGGAGG 285

     127 yGlyPhe.HisGluArgGlyGln...AspTrpGluGlnThrGlnLysIle 142
     286 TAACTTCCTACCCACAAGGCTCTATCCCTGAAGAGGATGCTGGGAAATTA 335

     143 TyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTy 159
     336 TATAACACCTGTGCTGTGTTTGGGCCTGATGGAACTTTACTAGCAAAGTA 385

     159 rArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetC 176
     386 TAGAAAGATCCATCTGTTTGACATTGATGTTCCTGGAAAAATTACATTTC 435

     176 ysGluSerAsnSerThrMetProGlyProSerLeuGlu............ 188
     436 AAGAATCTAAAACATTGAGTCCGGGTGATAGTTTCTCCACATTTGATACT 485

     188 ..................................................... 188

     486 CGTATGTACCAGATAAGTTTGCCTCTTTAGCAATCTCAGTAGAAGACAAT 535

     189 .......................SerProValSer............. 192
     536 CAGGTATTTATTTCTTTTTTGTCTCTCTCCGATTTCTTCACATAACCTAA 585

     192 ..................................................... 192

     586 CTGAAAGACCATAAGTGAGAAAGGCAGAGAATCATCACAGATCTGGAAAG 635

     192 ..................................................... 192

     636 TTCGGGCTTATTTGAGAACTAAGGATTTGACACGATTTTGCCCTTTGATT 685

     192 ..................................................... 192

     686 TGATTGTAGCTTCCTGTTACGGCTTCCAGAGTATACCTATTAGGCTACAG 735

     192 ..................................................... 192

     736 TTGAGTACCTCCCATCTAGATAATAAGCATTCAATTAGAATGAATTTCTC 785

     193 ......ThrProAla.GlyLys.......................... 197
     786 ATCTTTACTCCGCTGATGTAAATGATGTCTTTATGAGATGAAGTCCAAGT 835

     197 ..................................................... 197

     836 AGGAATGAGCTTGTAAATTATCTCTGTCCTCAGGTCCTGTGTTAATTTAT 885

     197 ..................................................... 197

     886 CCCTGTCAGTGTTTTGTGATCATTATGTCATGGAGGATTTCCCCTGCCAC 935

     197 ..................................................... 197

     936 ACCATGCTGTAGGGAGTTAACTTTTCATTTGTGCATTTTCTGTTTGGAAA 985

     198 ............IleGlyLeuAlaValCysTyrAspMetArgPhePro 209
     986 CAGCTTACTGCAGAGTGGGTCTGGGCATCTGCTACGACATGCGGTTTGCA 1035

     210 GluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPr 226
    1036 GAGCTTGCACAAATCTACGCACAGAGAGGCTGCCAGCTGTTGGTATATCC 1085

     226 oSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuA 243
    1086 AGGAGCTTTTAATCTGACCACTGGACCAGCCCATTGGGAGTTACTTCAGC 1135

     243 rgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 259
    1136 GAAGCCGGGCTGTTGATAATCAGGTGTATGTGGCCACAGCCTCTCCTGCC 1185

     260 GlyArgHisHisGluLysArgAlaSerTyr.........GlyHisSerMe 273
    1186 .........CGGGATGACAAAGCCTCCTATGTTGCCTGGGGACACAGCAC 1226

     273 tValValAspProTrpGlyThrValValAlaArgCysSerGluGlyProG 290
    1227 CGTGGTGAACCCTTGGGGGGAGGTTCTAGCCAAAGCTGGCACAGAAGAAG 1276

     290 lyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArg 306
    1277 CAATCGTGTATTCAGACATAGACCTGAAGAAGCTGGCTGAAATACGCCAG 1326

     307 HisLeuProValPheGlnHisArgArgProAspLeuTyr 319
    1327 CAAATCCCCGTTTTTAGACAGAAGCGATCAGACCTCTAT 1365

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL13787

seq_documentation_block:
ID   ABL13787 standard; cDNA; 898 BP.
XX
AC   ABL13787;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 35843.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
```

```
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69684.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 35843; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 898 BP; 243 A; 236 C; 231 G; 188 T; 0 other;


alignment_scores:
           Quality:  387.00            Length:    293
             Ratio:  2.138               Gaps:      5
 Percent Similarity:  61.775   Percent Identity: 31.399

alignment_block:
 US-09-357-675C-21 x ABL13787   ..

 Align seg 1/1  to: ABL13787  from: 1  to: 898

      41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerTh 57
         |||::::::   ::::::     :::|||:::   |||:::   :::::
      50 AGCAAAACTAGCAACATCATGCGTTTGGCGCTGCTGCAGCTCAAGGGTTC 99

      57 rProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluA 74
         :   ||||||      |||   ::::::::::::      ::::::   |
     100 CAAGGACAAAGTGGCCAATGTCCAAAACGCCGTCACCAAAATCGAGGCGG 149

      74 laAlaArg...LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAsp 89
         ||::::::               |||      |||||||||:::|||
     150 CGGTCAAGGAGCATAAACCCCGATTGATCACTCTGCCGGAGTGCTTT... 196

      90 PheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGl 106
                                          :::   |||   ||
     197 ..................................AATGCTCCATACGG 210

     106 yGlyLysLeuLeuGluGluTyrThrGln..................... 115
         |   |||   ::::::||||||::::::
     211 CACCAAGTACTTTCGGGAGTACTCCGAGACCATTCCCGATGGCTATACCT 260

     116 .................LeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
                         |||||||||:::      :::::::::
     261 CGCAACAGCTCTCCAATTTGGCCAGGAAGCACCAGGTGTACATCGTGGGC 310

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
         |||   :::   |||   |||::::::            :::   |||||
     311 GGAACAATCCCTGAACTGGGCGAAAAC.............GATGCCATCTA 348

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
         ||||      ::::::   :::      |||   :::||||||   :::|
     349 CAACACCTGCACGGTTTGGTCACCCACTGGTGACCTTGTGGCCAAGCATC 398

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
         |||||   ||||||   |||:::::::::   |||   |||
     399 GCAAGATGCATCTCTTTGACATAGATGTCAAGGGT...GGCATTCGCTTC 445

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
         ::::::::::::|||:::   :::   :::      :::   ::::::::
     446 AAGGAGTCGGAAACGCTGTCCGCAGGCAATGATTTCACCATCATCAACGT 495

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
         :   :::   |||||||||:::::::::|||||||||:::||||||   |
     496 AGATGGCCACAAGATTGGCATCGGCATCTGCTACGATATTCGATTCGAGG 545

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
         ||:::::            :::||||||:::|||::::::   ||||||
     546 AGATGGCGAGGCTCTATCGCAACGCAGGCTGCGAGATGATCATCTATCCG 595

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
         :::||||||:::      |||||||||   |||||||||:::|||   ||
     596 GCTGCATTCAACATGACCACTGGTCCACTGCACTGGGAGCTATTGCAGCG 645

     243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
         |:::||||||   ::::::|||   :::||||||:::::::::   :::
     646 ATCCCGTGCCAATGACAACCAACTCTTTGTGGTCACAACATCACCAGCCC 695

     260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
                       :::   ::::::||||||||||||||||||||:::
     696 GTGACACAAGCGCCGAGTATGTAGCCTATGGCCATTCCATGGTGGTGAAT 745

     277 ProTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLe 293
         ||||||:::   |||         :::|||||||||      :::   ::
     746 CCATGGGCCAAGGTGCAGCAGAGTGCCAGTGAAGGCGAGGAAATCGTGGT 795

     293 uAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProV 310
         :|||   ||||||::::::   ::::::|||:::|||:::::::::||||
     796 GGCCGATATAGATTTCTCCGAGGTGGAGCAGGTGCGTCAGCAGATTCCCG 845

     310 alPheGlnHisArgArgProAspLeuTyr 319
         |||||   :::||||||   ||||||||
     846 TCTTTGGGCAAAGACGTCTAGATCTGTAC 874
```

```
OM of: US-09-357-675C-21 to: Issued_Patents_NA:*   out_format : pfs

Date: Apr 29, 2002 10:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091757_3210/app_query.fasta_1.1208
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09357675_@CGN1_1_118 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 85.140000


score_list:
Sequence                Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-63 +  206.00  364.19  1.3e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-63 +  206.00  364.19  1.3e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-25 +  205.00  362.25  1.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-25 +  205.00  362.25  1.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-5 +  204.00  360.31  2.2e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-59 +  204.00  360.31  2.2e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-5 +  204.00  360.31  2.2e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-59 +  204.00  360.31  2.2e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-61 +  203.00  358.37  2.8e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-61 +  203.00  358.37  2.8e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-9 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-11 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-21 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-11 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-21 +  202.00  356.43  3.6e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-917-111-1 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-479-638-1 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-1 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-3 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-7 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-13 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-15 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-17 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-19 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-23 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-27 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-29 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-31 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-47 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-49 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-51 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-55 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-65 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-1 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-3 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-7 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-13 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-15 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-17 +  201.00  354.49  4.7e-12   1785  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-19 +  201.00  354.49  4.7e-12   1785
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-23 +  201.00  354.49  4.7e-12   1785
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-27 +  201.00  354.49  4.7e-12   1785
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-29 +  201.00  354.49  4.7e-12   1785
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-31 +  201.00  354.49  4.7e-12   1785

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-63

seq_documentation_block:
; Sequence 63, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD469
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-63


alignment_scores:
            Quality:  206.00            Length:    340
              Ratio:  1.280               Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-63   ..

 Align seg 1/1  to: US-08-294-871A-63  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     ::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
          :::::::::   :::||||||      ||||||  :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||               |||:::   |||             |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
          ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::       :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
          |||||||||   ||||||               ||||||:::

     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
             |||   ::::::      |||          ||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.............................AlaGly 196
             |||                                 |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::      ||||||||||||:::                  |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
         ||:::   :::   |||:::|||      ::::::
     810 GCGGCTACAACACGCCGACCCACAATCCCATTGTTCCCCAGCACGACCAC 859

     242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                  ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||::::::   |||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                 |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::                :::      |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-63

seq_documentation_block:
; Sequence 63, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD469
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-63


alignment_scores:
            Quality:  206.00              Length:    340
              Ratio:  1.280                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-63   ..

 Align seg 1/1  to: US-08-876-398A-63  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                      ::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
            :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||             |||:::   |||             |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
          ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||                :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
            |||||||||   ||||||               |||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
                |||   ::::::      |||         ||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro...............................AlaGly 196
                 |||                                |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::      ||||||||||||:::                  |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
         ||:::   :::   |||:::|||      ::::::
     810 GCGGCTACAACACGCCGACCCACAATCCCATTGTTCCCCAGCACGACCAC 859

     242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                  ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||::::::   |||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                 |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::               :::      |||   :::||||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-25

seq_documentation_block:
; Sequence 25, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
```

```
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD445
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-25


alignment_scores:
            Quality:  205.00              Length:    335
              Ratio:  1.273                 Gaps:     15
 Percent Similarity:  48.060   Percent Identity: 25.970

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-25   ..

 Align seg 1/1  to: US-08-294-871A-25  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                      :::::::|||
     164 GCGATGTGTTTCACAACGTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         :::::::::   :::||||||      ||||||   :::       ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||              |||:::   |||           |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
          ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
          |||||||||   ||||||                |||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
             |||   ::::::      |||         |||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.............................AlaGly 196
             |||                                 |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr......TyrProSerA 228
          :::::::::      ||||||||||||:::         |||   :::
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGCGGCGGCTACAACACGC 824

     228 laPheGlySerIleThrGlyProAlaHisTrpGluValLeu......... 241
```

```
                :::   |||   |||   |||   ::::::
    825 CGACCCACAATCCCACCGTTCCCCAGCACGACCACCTGACGTCCTTCCAC 874

    242 ............LeuArgAlaArgAlaIleGluThrGlnCysTyrValVa 254
                    ::::::|||   :::   ::::::   ::::::   ::
    875 CATCTCCTATCGATGCAGGCCGGGTCTTATCAGAACGGGGCCTGGTCCGC 924

    254 lAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyH 271
        :||||||::::::   |||      :::|||:::           ||||
    925 GGCCGCGGGCAAGGTGGGCATG...GAGGAGAACTGCATGCTGCTCGGCC 971

    271 isSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 287
        |||||   :::|||   |||   |||   :::||||||      :::
    972 ACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTCGCTCTCACTACGACG 1021

    288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGl 303
                :::      |||   :::||||||:::      |||::
   1022 CTGGAAGACGAGGTGATCACCGCCGCCGTCGATCTCGATCGCTGCCGGGA 1071

    303 nLeuArgArgHisLeuProValPhe...GlnHisArgArgProAspLeuT 319
        :||||||:::|||:::      |||   |||||||||:::|||:::   |
   1072 ACTGCGTGAACACATCTTCAACTTCAAGCAGCATCGTCAGCCCCAGCACT 1121

    319 yrGly 320
        |||||
   1122 ATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-25

seq_documentation_block:
; Sequence 25, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD445
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-25


alignment_scores:
            Quality:  205.00              Length:    335
              Ratio:  1.273                Gaps:     15
 Percent Similarity:  48.060   Percent Identity: 25.970

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-25   ..

 Align seg 1/1  to: US-08-876-398A-25  from: 1  to: 1785

     50 AlaValCysGlnValThr.....................SerThrProAs 59
        |||:::|||   :::|||                     ::::::|||
    164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

     59 pLysGlnGlnAsnPheLysThr............................ 66
        |||:::|||:::|||   |||
    214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

     67 .............................................Cys 67

    264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

     68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
           :::::::::   :::|||||||      ||||||   :::      ::
    314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

     84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
        :||||||               |||:::   |||            |||
    364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

     95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
          ||||||        :::   |||      :::   |||   ::::::
    414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

    111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
        :::         :::   |||   |||   |||:::   ::::::|||||
    464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

    127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
        |      :::         |||   |||                :::::::
    514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

    141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
        ::   :::|||      :::|||:::::::::::::|||   :::|||:::
    549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

    158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||              ||||||:::
    599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

    174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
              |||   ::::::      |||       ||||||:::
    632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674
```

```
    191 alSerThrPro.............................AlaGly 196
                |||                                 |||:::
    675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

    197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
        |||:::|||:::   :::|||   |||   |||:::||||||
    725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

    213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr......TyrProSerA 228
         :::::::::      ||||||||||||:::         |||   :::
    775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGCGGCGGCTACAACACGC 824

    228 laPheGlySerIleThrGlyProAlaHisTrpGluValLeu......... 241
                :::   |||   |||   |||   ::::::
    825 CGACCCACAATCCCACCGTTCCCCAGCACGACCACCTGACGTCCTTCCAC 874

    242 ............LeuArgAlaArgAlaIleGluThrGlnCysTyrValVa 254
                   ::::::|||   :::   ::::::   ::::::   ::
    875 CATCTCCTATCGATGCAGGCCGGGTCTTATCAGAACGGGGCCTGGTCCGC 924

    254 lAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyH 271
        :||||||::::::   |||      :::|||:::            ||||
    925 GGCCGCGGGCAAGGTGGGCATG...GAGGAGAACTGCATGCTGCTCGGCC 971

    271 isSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 287
        |||||   :::|||   |||   |||   :::||||||      :::
    972 ACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTCGCTCTCACTACGACG 1021

    288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGl 303
                  :::      |||   :::||||||:::      |||::
   1022 CTGGAAGACGAGGTGATCACCGCCGCCGTCGATCTCGATCGCTGCCGGGA 1071

    303 nLeuArgArgHisLeuProValPhe...GlnHisArgArgProAspLeuT 319
        :||||||:::|||:::      |||   ||||||||:::|||:::   |
   1072 ACTGCGTGAACACATCTTCAACTTCAAGCAGCATCGTCAGCCCCAGCACT 1121

    319 yrGly 320
        |||||
   1122 ATGGT 1126
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-5

seq_documentation_block:
; Sequence 5, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD404 (FERM BP-3913)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-5


alignment_scores:
            Quality:  204.00              Length:    340
              Ratio:  1.267                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-5   ..

 Align seg 1/1  to: US-08-294-871A-5  from: 1  to: 1785

      50 AlaValCysGlnValThr...................SerThrProAs 59
         |||:::|||   :::|||                   :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213
```

```
 59 pLysGlnGlnAsnPheLysThr..............................  66
    |||:::|||:::|||   |||
214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

 67 .............................................Cys  67

264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

 68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe  84
     :::::::::   :::|||||||      ||||||   :::      ::
314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

 84 uProGlu......AlaPheAspPheIleAlaArg............AspP  95
    :||||||               |||::: ' |||            |||
364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

 95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
      ||||||         :::   |||      :::   |||   ::::::
414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
    :::         :::   |||   |||   |||:::   ::::::|||||
464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
    |       :::         |||   |||                :::::::
514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
    ::   :::|||      :::|||:::::::::::::|||   :::|||:::
549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
      |||||||||   ||||||               ||||||:::
599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
           |||   ::::::      |||          ||||||:::
632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

191 alSerThrPro.................................AlaGly 196
            |||                                ||||::
675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
    |||:::|||:::   :::|||   |||   |||:::||||||
725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
     :::::::::      ||||||||||||:::                  |
775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
    ||:::   :::   |||:::|||      ::::::
810 GCGGCTACAACACGCCGACCCACAATCCCCTTGTTCCCCAGCACGACCAC 859

242 ........................LeuArgAlaArgAlaIleGluTh 249
                            ::::::|||   :::   ::::::
860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
    :   ::::::   :::||||||::::::   |||      :::|||:::
910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
          |||||||||   :::|||   |||   |||   :::|||
957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
    |||      :::               :::      |||   :::||||||
1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
    |:::      |||:::||||||:::|||:::      |||   |||||||
1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

314 rgArgProAspLeuTyrGly 320
    ||:::|||:::   ||||||
1107 GTCAGCCCCAGCACTATGGT 1126
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-59

seq_documentation_block:
; Sequence 59, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  PCT/JP93/01101
;       FILING DATE:  05-AUG-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 340078/1992
;       FILING DATE:  21-DEC-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wegner, Harold C.
;       REGISTRATION NUMBER:  25,258
;       REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 672-5300
;       TELEFAX:  (202) 672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1785 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       STRAIN:  JM109 pAD456
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  233..1144
US-08-294-871A-59


alignment_scores:
            Quality:  204.00              Length:    340
              Ratio:  1.267                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-59   ..

 Align seg 1/1  to: US-08-294-871A-59  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .................................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
            :::::::::   :::||||||      ||||||   :::       ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||              |||:::  |||              |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGTATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
            |||||||||   ||||||               |||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
            |||   ::::::      |||          ||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
               |||                                   |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::      ||||||||||||:::                  |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
         ||:::   :::   |||:::|||      ::::::
     810 GCGGCTACAACACGCCGACCCACAATCCCCTTGTTCCCCAGCACGACCAC 859

     242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                      ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||::::::   |||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGTCAGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                  |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::              :::      |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::     |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-5

seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/876,398A
;       FILING DATE:  16-JUN-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/211,641
;       FILING DATE:  11-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/JP93/01101
;       FILING DATE:  05-AUG-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 340078/1992
;       FILING DATE:  21-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 212692/1992
;       FILING DATE:  10-AUG-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wegner, Harold C.
;       REGISTRATION NUMBER:  25,258
;       REFERENCE/DOCKET NUMBER:  74129/130
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 672-5300
;       TELEFAX:  (202) 672-5399
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1785 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli
;       STRAIN:  JM109 pAD404 (FERM BP-3913)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(233..1141)
US-08-876-398A-5


alignment_scores:
            Quality:  204.00              Length:    340
              Ratio:  1.267                Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-5   ..

 Align seg 1/1  to: US-08-876-398A-5  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
            :::::::::   :::||||||      ||||||   :::       ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||               |||:::   |||           |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||  . ::::::

     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::       :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |       :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
             |||||||||   ||||||              ||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
               |||   ::::::      |||         ||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
              |||                                   |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::      ||||||||||||:::                   |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
         ||:::   :::   |||:::|||      ::::::
     810 GCGGCTACAACACGCCGACCCACAATCCCCTTGTTCCCCAGCACGACCAC 859

     242 ........................LeuArgAlaArgAlaIleGluTh 249
                                 :::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||:::::   |||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                 |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||       :::                 :::       |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-59

seq_documentation_block:
; Sequence 59, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;     APPLICANT:  IKENAKA, Yasuhiro
;     APPLICANT:  NANBA, Hirokazu
;     APPLICANT:  TAKANO, Masayuki
;     APPLICANT:  YAJIMA, Kazuyoshi
;     APPLICANT:  YAMADA, Yukio
```

```
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD456
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-59


alignment_scores:
           Quality:  204.00            Length:    340
             Ratio:  1.267               Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-59   ..

 Align seg 1/1  to: US-08-876-398A-59  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr.......................... 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263
```

```
      67 ..............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
            :::::::::   :::||||||      ||||||   :::       ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg.............AspP 95
         :||||||                 |||:::   |||           |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGTATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   :::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||                :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||       :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
            |||||||||   ||||||                |||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
               |||   ::::::   .  |||         ||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
               |||                                   |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::       ||||||||||||:::                  |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
         |||:::   :::   |||:::|||      ::::::
     810 GCGGCTACAACACGCCGACCCACAATCCCCTTGTTCCCCAGCACGACCAC 859

     242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                  ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||::::::   |||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGTCAGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                  |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::              :::       |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
```

```
          ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-61

seq_documentation_block:
; Sequence 61, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD468
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-61


alignment_scores:
           Quality:  203.00              Length:    340
             Ratio:  1.261                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-61   ..

 Align seg 1/1  to: US-08-294-871A-61  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     ::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||               |||:::   |||            |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||               ||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
```

```
            |||   ::::::      |||         ||||||:::
    632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

    191 alSerThrPro................................AlaGly 196
                |||                                 |||:::
    675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

    197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
        |||:::|||:::   :::|||   |||   |||:::||||||
    725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

    213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
         :::::::::      ||||||||||||:::                  |
    775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

    230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
        ||:::   :::   |||:::|||      ::::::
    810 GCGGCTACAACACGCCGACCCACAATCCCGCTGTTCCCCAGCACGACCAC 859

    242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                  :::::::|||   :::   :::::
    860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

    249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
        :   ::::::   :::||||||::::::   |||      :::|||:::
    910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

    266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                |||||||||   :::|||   |||   |||   :::|||
    957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

    283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
        |||      :::                :::       |||   :::|||||
   1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

    298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
        |:::      |||:::|||||||:::|||:::      |||   |||||||
   1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

    314 rgArgProAspLeuTyrGly 320
        ||:::|||:::   ||||||
   1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-61

seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE: . 05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD468
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-61


alignment_scores:
            Quality:  203.00            Length:    340
              Ratio:  1.261               Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-61   ..

 Align seg 1/1  to: US-08-876-398A-61  from: 1  to: 1785

     50 AlaValCysGlnValThr.....................SerThrProAs 59
        |||:::|||   :::|||                     :::::::|||
    164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

     59 pLysGlnGlnAsnPheLysThr............................ 66
         |||:::|||:::|||   |||
    214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

     67 .............................................Cys 67

    264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

     68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
           :::::::::   :::||||||     ||||||   :::       ::
    314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

     84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
        :||||||              |||:::   |||            |||
    364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

     95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
         ||||||         :::   |||       :::   |||   ::::::
    414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

    111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
        :::       :::   |||   |||  |||:::   ::::::|||||
    464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513
```

```
    127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
        |      :::         |||   |||                :::::::
    514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

    141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
        ::   :::|||      :::|||:::::::::::::|||   :::|||::: 
    549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

    158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||              ||||||::: 
    599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

    174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
            |||   ::::::      |||         |||||||::: 
    632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

    191 alSerThrPro.................................AlaGly 196
                |||                                 |||::: 
    675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

    197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
        |||:::|||:::   :::|||   |||   |||:::||||||       
    725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

    213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
         :::::::::      ||||||||||||:::                  | 
    775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

    230 lySerIleThrGlyProAlaHisTrpGluValLeu.............. 241
        ||:::   :::   |||:::|||      ::::::
    810 GCGGCTACAACACGCCGACCCACAATCCCGCTGTTCCCCAGCACGACCAC 859

    242 ..........................LeuArgAlaArgAlaIleGluTh 249
                                 :::::::|||   :::   ::::: 
    860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

    249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
        :   ::::::   :::||||||::::::   |||      :::|||::: 
    910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGTGGGCATG...GAGGAGAACT 956

    266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                |||||||||   :::|||   |||   |||   :::||| 
    957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

    283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
        |||      :::                :::      |||   ::::|||||
   1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

    298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
        |:::      |||:::||||||:::|||:::      |||   ||||||| 
   1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

    314 rgArgProAspLeuTyrGly 320
        ||:::|||:::   ||||||
   1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-9

seq_documentation_block:
; Sequence 9, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD416 (FERM BP-3915)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
```

```
US-08-294-871A-9


alignment_scores:
            Quality:  202.00              Length:    340
              Ratio:  1.255                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-9   ..

 Align seg 1/1  to: US-08-294-871A-9  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                      :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||              |||:::   |||            |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||                :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||               |||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
               |||   ::::::     |||        |||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
               |||                                   |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
           :::::::::      ||||||||||||:::                 |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAla............................ 236
         ||:::   :::   |||:::
     810 GCGGCTACAACACGCCGACCCACAATCCCCCTGTTCCCCAGCACGACCAC 859

     237 ............HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
                       |||       :::   :::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||:::::::::|||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGCGGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                   |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::               :::      |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-11

seq_documentation_block:
; Sequence 11, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD428
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-11


alignment_scores:
            Quality:  202.00              Length:    340
              Ratio:  1.255                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-11   ..

 Align seg 1/1  to: US-08-294-871A-11  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     :::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr............................ 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
          :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||              |||:::   |||            |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::          |||   |||              :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||               ||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
               |||   ::::::     |||        |||||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
                 |||                                  |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
          :::::::::       ||||||||||||:::                 |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAla.............................. 236
         ||:::   :::   |||:::
     810 GCGGCTACAACACGCCGACCCACAATCCCCCTGTTCCCCAGCACGACCAC 859

     237 ............HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
                     |||      :::   ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||:::::::::|||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGCCGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                 |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::                :::      |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-21

seq_documentation_block:
; Sequence 21, Application US/08294871A
; Patent No. 5824522
;  GENERAL INFORMATION:
;    APPLICANT:  Ikenaka, Yasuhiro
;    APPLICANT:  Nanba, Hirokazu
;    APPLICANT:  Takano, Masayuki
```

```
;    APPLICANT:  Yajima, Kazuyoshi
;    APPLICANT:  Yamada, Yukio
;    APPLICANT:  Takahashi, Satomi
;    APPLICANT:  Okubo, Kazuma
;    APPLICANT:  Yamada, Kazuhiko
;    APPLICANT:  Hiraishi, Yoshiro
;    TITLE OF INVENTION:  Immobilized Enzyme Preparation and
;    TITLE OF INVENTION:  Process for Producing D-a-Amino Acid
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,871A
;      FILING DATE:  22-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,758
;      FILING DATE:  12-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,111
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 400848/1990
;      FILING DATE:  07-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP91/01696
;      FILING DATE:  06-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 407922/1990
;      FILING DATE:  27-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 078840/1991
;      FILING DATE:  11-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 140051/1991
;      FILING DATE:  12-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP92/00739
;      FILING DATE:  10-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/127/AOPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  JM109 pAD439
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  233..1144
US-08-294-871A-21


alignment_scores:
          Quality:  202.00            Length:    340
            Ratio:  1.255               Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-21   ..

 Align seg 1/1  to: US-08-294-871A-21  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                      :::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr.......................... 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 ..............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
         :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||               |||:::   |||            |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::         |||   |||               :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
            |||||||||   ||||||               ||||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
                |||   ::::::      |||         |||||:::
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.................................AlaGly 196
                 |||                                |||:::
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
         |||:::|||:::   :::|||   |||   |||:::||||||
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
```

```
          :::::::::      ||||||||||||:::                  |
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

     230 lySerIleThrGlyProAla............................ 236
         ||:::   :::   |||:::
     810 GCGGCTACAACACGCCGACCCACAATCCCCCTGTTCCCCAGCACGACCAC 859

     237 ............HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
                    |||      :::   ::::::|||   :::   :::::
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :   ::::::   :::||||||:::::::::|||      :::|||:::
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGCTGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                 |||||||||   :::|||   |||   |||   :::|||
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
         |||      :::                :::      |||   :::|||||
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
         |:::      |||:::||||||:::|||:::      |||   |||||||
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
         ||:::|||:::   ||||||
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9

seq_documentation_block:
; Sequence 9, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD416 (FERM BP-3915)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-9


alignment_scores:
            Quality:  202.00            Length:    340
              Ratio:  1.255               Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-9   ..

 Align seg 1/1  to: US-08-876-398A-9  from: 1  to: 1785

      50 AlaValCysGlnValThr.....................SerThrProAs 59
         |||:::|||   :::|||                     ::::::|||
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr........................... 66
          |||:::|||:::|||   |||
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
           :::::::::   :::||||||      ||||||   :::      ::
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
         :||||||         :::   |||      :::   |||         |||
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
           ||||||         :::   |||      :::   |||   ::::::
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
         :::         :::   |||   |||   |||:::   ::::::|||||
     464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

     127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
         |      :::          |||   |||                :::::::
     514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

     141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
         ::   :::|||      :::|||:::::::::::|||   :::|||:::
     549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

     158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
           |||||||||   ||||||                |||||:::
     599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631
```

```
     174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
     632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

     191 alSerThrPro.............................AlaGly 196
     675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

     197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
     725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

     213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
     775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC.............G 809

     230 lySerIleThrGlyProAla..................................... 236
     810 GCGGCTACAACACGCCGACCCACAATCCCCCTGTTCCCCAGCACGACCAC 859

     237 ............HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
     860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
     910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGCGGGCATG...GAGGAGAACT 956

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
     957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

     283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
    1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

     298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
    1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

     314 rgArgProAspLeuTyrGly 320
    1107 GTCAGCCCCAGCACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-11

seq_documentation_block:
; Sequence 11, Application US/08876398A
; Patent No. 6083752
;  GENERAL INFORMATION:
;    APPLICANT:  IKENAKA, Yasuhiro
;    APPLICANT:  NANBA, Hirokazu
;    APPLICANT:  TAKANO, Masayuki
;    APPLICANT:  YAJIMA, Kazuyoshi
;    APPLICANT:  YAMADA, Yukio
;    APPLICANT:  TAKAHASHI, Satomi
;    TITLE OF INVENTION:  DNA CODING FOR DECARBAMYLASE IMPROVED IN
;    TITLE OF INVENTION:  THERMOSTABILITY AND USE THEREOF
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,398A
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/211,641
;      FILING DATE:  11-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP93/01101
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 340078/1992
;      FILING DATE:  21-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 212692/1992
;      FILING DATE:  10-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wegner, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  74129/130
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  JM109 pAD428
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(233..1141)
US-08-876-398A-11


alignment_scores:
            Quality:  202.00              Length:    340
              Ratio:  1.255                 Gaps:     15
 Percent Similarity:  47.353   Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-876-398A-11   ..

 Align seg 1/1  to: US-08-876-398A-11  from: 1  to: 1785

      50 AlaValCysGlnValThr.....:.....SerThrProAs 59
     164 GCGATGTGTTTCACAACGTTTTCCCGGCCGCTGGGCCGGACATCACCTAG 213

      59 pLysGlnGlnAsnPheLysThr.......................... 66
     214 GAAGGAGCAGAGGTTCATGACACGTCAGATGATACTTGCAGTGGGACAAC 263

      67 .............................................Cys 67

     264 AAGGTCCGATCGCGCGCGCGGAGACACGCGAACAGGTCGTCGTTCGTCTT 313

      68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
     314 CTCGACATGCTGACGAAAGCCGCGAGCCGGGGCGCGAATTTCATTGTCTT 363

      84 uProGlu......AlaPheAspPheIleAlaArg............AspP 95
     364 CCCCGAACTCGCGCTTACGACCTTCTTCCCGCGCTGGCATTTCACCGACG 413

      95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
     414 AGGCCGAGCTCGATAGCTTCTATGAGACCGAAATGCCCGGCCCGGTGGTC 463

     111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
```

```
 464 CGTCCACTCTTTGAGAAGGCCGCGGAACTCGGGATCGGCTTCAATCTGGG 513

 127 yGlyPheHis.........GluArgGlyGlnAspTrpGluGlnThrGlnL 141
 514 CTACGCTGAACTCGTCGTCGAAGGCGGC..............GTCAAGC 548

 141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
 549 GTCGCTTCAACACGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598

 158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
 599 AAGTATCGTAAGATCCATTTG..............CCGGGTCACAAG.. 631

 174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
 632 .......GAGTACGAGGCCTACCGGCCGTTCCAGCATCTTGAAAAGCGTT 674

 191 alSerThrPro.............................AlaGly 196
 675 ATTTCGAGCCGGGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCCGCG 724

 197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
 725 AAAATGGGGATGTTCATCTGCAACGATCGCCGCTGGCCTGAAGCCTGGCG 774

 213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
 775 GGTGATGGGCCTCAGGGGCGCCGAGATCATCTGC..............G 809

 230 lySerIleThrGlyProAla.............................. 236
 810 GCGGCTACAACACGCCGACCCACAATCCCCCTGTTCCCCAGCACGACCAC 859

 237 ............HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
 860 CTGACGTCCTTCCACCATCTCCTATCGATGCAGGCCGGGTCTTATCAGAA 909

 249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
 910 CGGGGCCTGGTCCGCGGCCGCGGGCAAGGCCGGCATG...GAGGAGAACT 956

 266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
 957 GCATGCTGCTCGGCCACTCCTGCATCGTGGCGCCGACCGGGGAAATCGTC 1006

 283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCCGCCGTCGATCT 1056

 298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
1057 CGATCGCTGCCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106

 314 rgArgProAspLeuTyrGly 320
1107 GTCAGCCCCAGCACTATGGT 1126
```

```
OM of: US-09-357-675C-21 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Apr 29, 2002 12:18 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1h
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091758_3302/app_query.fasta_1.1208
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=p2n.rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09357675_@CGN1_1_10416 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: Pending_Patents_NA_Main:*
Database sequences: 21979536
Database length: -2067452561
Search time (sec): 4349.110000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/pna/US096C_COMB.seq:US-09-652-125A-9045 + 1748.00 2640.38  1.3e-137  1
/cgn2_6/ptodata/2/pna/US097B_COMB.seq:US-09-726-802-2166 + 1748.00 2640.38  1.3e-137  18
/cgn2_6/ptodata/2/pna/US6021_COMB.seq:US-60-213-360-100 + 1748.00 2632.52  3.5e-137  417
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-357-675A-1 + 1743.00 2635.19  2.5e-137  1416
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-357-675C-1 + 1743.00 2635.19  2.5e-137  1416
/cgn2_6/ptodata/2/pna/US6027_COMB.seq:US-60-278-258-200 - 1730.00 2607.89  8.3e-136  313
/cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:PCT-US00-05988-692 + 1718.00 2597.42  3.2e-135  138
/cgn2_6/ptodata/2/pna/US099A_COMB.seq:US-09-925-300-692 + 1718.00 2597.42  3.2e-135  138
/cgn2_6/ptodata/2/pna/US6017_COMB.seq:US-60-172-373-14716 + 1718.00 2585.13  1.5e-134  5
/cgn2_6/ptodata/2/pna/US6016_COMB.seq:US-60-164-285-3819 + 1717.00 2585.99  1.4e-134  39
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-359-922-12822 + 1626.50 2448.90  5.9e-127  37
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-359-922-12822 + 1626.50 2448.90  5.9e-127  37
/cgn2_6/ptodata/2/pna/US6036_COMB.seq:US-60-360-207-3686 + 1487.00 2246.87  1.1e-115  12
/cgn2_6/ptodata/2/pna/US6024_COMB.seq:US-60-242-679-78 - 1468.50 2208.98  1.4e-113  3620
/cgn2_6/ptodata/2/pna/US6032_COMB.seq:US-60-324-185-10674 + 1315.00 1971.69  2.3e-100  5
/cgn2_6/ptodata/2/pna/US6024_COMB.seq:US-60-242-679-1588 + 1303.00 1972.36  2.1e-100  75
/cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:PCT-US98-16235-92 + 1067.00 1609.11  3.5e-80   1203
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-244-112-92 + 1067.00 1609.11  3.5e-80   1203
/cgn2_6/ptodata/2/pna/US097C_COMB.seq:US-09-774-639-92 + 1067.00 1609.11  3.5e-80   1203
/cgn2_6/ptodata/2/pna/US099C_COMB.seq:US-09-969-730-76 + 1067.00 1609.11  3.5e-80   1203
/cgn2_6/ptodata/2/pna/US095B_COMB.seq:US-09-535-896-18207 +  994.00 1504.83  2.3e-74   5
/cgn2_6/ptodata/2/pna/US098B_COMB.seq:US-09-855-294B-10 +  968.50 1459.29  7.8e-72   121
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-354-899-1617 +  873.00 1320.80  4.1e-64   601
/cgn2_6/ptodata/2/pna/US096C_COMB.seq:US-09-652-911-7920 +  845.50 1271.87  2.2e-61   12
/cgn2_6/ptodata/2/pna/US095B_COMB.seq:US-09-535-896-18219 +  801.00 1212.83  4.2e-58   5
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-235-076-23825 +  794.00 1202.82  1.5e-57   48
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-289-768-25986 +  794.00 1202.82  1.5e-57   48
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-332-782-23825 +  794.00 1202.82  1.5e-57   48
/cgn2_6/ptodata/2/pna/US097B_COMB.seq:US-09-737-223-23825 +  794.00 1202.82  1.5e-57   4
/cgn2_6/ptodata/2/pna/US099B_COMB.seq:US-09-939-397-25986 +  794.00 1202.82  1.5e-57   4
/cgn2_6/ptodata/2/pna/US098B_COMB.seq:US-09-834-366-11899 +  792.00 1198.45  2.6e-57   5
/cgn2_6/ptodata/2/pna/US6019_COMB.seq:US-60-197-873-11899 +  792.00 1198.45  2.6e-57   5
/cgn2_6/ptodata/2/pna/US096C_COMB.seq:US-09-652-911-9443 +  775.50 1163.12  2.5e-55   16
/cgn2_6/ptodata/2/pna/US097A_COMB.seq:US-09-710-282-3968 +  775.50 1163.12  2.5e-55   16
/cgn2_6/ptodata/2/pna/US097A_COMB.seq:US-09-716-472-5258 +  775.50 1163.12  2.5e-55   16
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-235-076-31257 +  768.50 1164.44  2.1e-55   46
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-289-768-39162 +  768.50 1164.44  2.1e-55   46
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-332-782-31257 +  768.50 1164.44  2.1e-55   46
/cgn2_6/ptodata/2/pna/US097B_COMB.seq:US-09-737-223-31257 +  768.50 1164.44  2.1e-55   4
/cgn2_6/ptodata/2/pna/US099B_COMB.seq:US-09-939-397-39162 +  768.50 1164.44  2.1e-55   4
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-240-371-6315 +  760.00 1151.36  1.1e-54
/cgn2_6/ptodata/2/pna/US092_COMB.seq:US-09-277-227-19834 +  760.00 1151.36  1.1e-54
/cgn2_6/ptodata/2/pna/US099A_COMB.seq:US-09-909-627-19834 +  760.00 1151.36  1.1e-54
/cgn2_6/ptodata/2/pna/US099A_COMB.seq:US-09-919-724-6315 +  760.00 1151.36  1.1e-54
/cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-306-609-11407 +  758.50 1147.16  1.9e-54

seq_name: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:US-09-652-125A-9045

seq_documentation_block:
; Sequence 9045, Application US/09652125A
; GENERAL INFORMATION:
;  APPLICANT: Shyjan, Andrew W.
;  APPLICANT:  DiStefano, Peter
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.1189-001
;  CURRENT APPLICATION NUMBER: US/09/652,125A
;  CURRENT FILING DATE:  2001-08-30
;  PRIOR APPLICATION NUMBER: 60/151,127
;  PRIOR FILING DATE: 1999-08-30
;  NUMBER OF SEQ ID NOS: 9506
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9045
;   LENGTH: 1826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-652-125A-9045


alignment_scores:
            Quality: 1748.00              Length:    327
              Ratio:  5.346                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x US-09-652-125A-9045   ..

 Align seg 1/1  to: US-09-652-125A-9045  from: 1  to: 1826

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     107 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 156

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 206

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 256

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 306

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 356

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 406

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 456

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 506

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     507 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 556

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 606

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 656

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     657 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 706

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     707 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 756

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     757 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 806

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     807 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 856

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     857 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 906

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     907 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 956

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     957 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 1006

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1007 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1056

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1057 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1087

seq_name: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:US-09-726-802-2166

seq_documentation_block:
; Sequence 2166, Application US/09726802
; GENERAL INFORMATION:
;  APPLICANT:  Acton, Susan L.
;  APPLICANT:  Falb, Dean A.
;  APPLICANT:  .
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.2025-001
;  CURRENT APPLICATION NUMBER: US/09/726,802
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER:  60/168,012
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  2872
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2166
;   LENGTH: 1826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-726-802-2166


alignment_scores:
           Quality: 1748.00            Length:    327
             Ratio:  5.346              Gaps:      0

 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x US-09-726-802-2166   ..

 Align seg 1/1  to: US-09-726-802-2166  from: 1  to: 1826

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     107 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 156

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 206

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 256

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 306

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 356

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 406

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 456

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 506

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     507 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 556

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 606

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 656

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     657 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 706

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     707 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 756

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     757 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 806

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     807 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 856

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     857 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 906

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     907 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 956

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     957 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 1006

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1007 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1056

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1057 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1087

seq_name: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:US-60-213-360-100

seq_documentation_block:
; Sequence 100, Application US/60213360
; GENERAL INFORMATION:
;  APPLICANT:  Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION:  Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide Polymor
;  TITLE OF INVENTION:  Identified Thereby
;  FILE REFERENCE:  GX-0014 P
;  CURRENT APPLICATION NUMBER: US/60/213,360
;  CURRENT FILING DATE:  2000-06-21
;  NUMBER OF SEQ ID NOS:  8347
;  SOFTWARE:  PERL Program
; SEQ ID NO 100
;   LENGTH: 4174
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 1096083.6c
;   NAME/KEY: unsure
;   LOCATION: 2331, 2333-2334, 2337, 2339-2340, 3582-3618
;   OTHER INFORMATION: a, t, c, g, or other
US-60-213-360-100


alignment_scores:
            Quality: 1748.00              Length:    327
              Ratio:  5.346                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x US-60-213-360-100   ..

 Align seg 1/1  to: US-60-213-360-100  from: 1  to: 4174

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 133

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 183

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 233

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 283

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 333

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 383

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 433

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     434 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 483

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     484 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 533

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     534 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 583

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     584 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 633

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     634 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 683

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     684 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 733

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     734 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 783

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     784 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 833

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     834 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 883

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     884 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 933

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     934 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 983

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     984 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1033

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1034 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1064

seq_name: /cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-357-675A-1

seq_documentation_block:
; Sequence 1, Application US/09357675A
; GENERAL INFORMATION:
;  APPLICANT:  Croce, M.D., Carlo M.
;  TITLE OF INVENTION:  Nitrilase Homologs
;  FILE REFERENCE:  CRO01.NP001
;  CURRENT APPLICATION NUMBER: US/09/357,675A
;  CURRENT FILING DATE:  1999-07-20
;  PRIOR APPLICATION NUMBER:  60/093,350
;  PRIOR FILING DATE:  1998-07-20
;  NUMBER OF SEQ ID NOS:  18
```

```
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1416
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: gene
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: n=a
US-09-357-675A-1


alignment_scores:
            Quality: 1743.00              Length:    326
              Ratio:  5.347                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x US-09-357-675A-1   ..

 Align seg 1/1  to: US-09-357-675A-1  from: 1  to: 1416

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 160

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCA 210

      35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 GAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTG 260

      52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGC 310

      68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     311 TGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360

      85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 410

     102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     411 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAG 460

     118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     461 GGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAG 510

     135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     511 ACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGC 560

     152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     561 AAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 610

     168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     611 GATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660

     185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 710

     202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     711 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGC 760

     218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     761 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCC 810

     235 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     811 CAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGC 860

     252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     861 TATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 910

     268 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     911 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

     285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTG 1010

     302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1011 CGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCT 1060

     318 uTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||
    1061 CTATGGCAATCTGGGTCACCCACTGTCT 1088

seq_name: /cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-357-675C-1

seq_documentation_block:
; Sequence 1, Application US/09357675C
; GENERAL INFORMATION:
;  APPLICANT: Croce, Carlo M.
;  TITLE OF INVENTION: Nitrilase Homologs
;  FILE REFERENCE: CRO01.NP001
;  CURRENT APPLICATION NUMBER: US/09/357,675C
;  CURRENT FILING DATE:  1999-07-20
;  PRIOR APPLICATION NUMBER: 60/093,350
;  PRIOR FILING DATE: 1998-07-20
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1416
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19)...(19)
;   OTHER INFORMATION: n=a
US-09-357-675C-1


alignment_scores:
            Quality: 1743.00              Length:    326
              Ratio:  5.347                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x US-09-357-675C-1   ..

 Align seg 1/1  to: US-09-357-675C-1  from: 1  to: 1416

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 160

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCA 210

      35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 211 GAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTG 260

  52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 261 TGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGC 310

  68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 311 TGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 360

  85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 361 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 410

 102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 411 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAG 460

 118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 461 GGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAG 510

 135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 511 ACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGC 560

 152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 561 AAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 610

 168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 611 GATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 660

 185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 661 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 710

 202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 711 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGC 760

 218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 761 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCC 810

 235 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 811 CAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGC 860

 252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 861 TATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAG 910

 268 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 911 TTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

 285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 961 GCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTG 1010

 302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1011 CGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCT 1060

 318 uTyrGlyAsnLeuGlyHisProLeuSer 327
     |||||||||||||||||||||||||||
1061 CTATGGCAATCTGGGTCACCCACTGTCT 1088

seq_name: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:US-60-278-258-200
```

```
seq_documentation_block:
; Sequence 200, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms Identified Thereby
;  FILE REFERENCE: GX-0010-1 P
;  CURRENT APPLICATION NUMBER: US/60/278,258
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 17730
;  SOFTWARE: PERL Program
; SEQ ID NO 200
;   LENGTH: 3130
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 1096083.6
;   NAME/KEY: unsure
;   LOCATION: 780-781, 783, 786-787, 789, 1508, 1550, 1570, 1574, 1722
;   OTHER INFORMATION: a, t, c, g, or other
US-60-278-258-200


alignment_scores:
            Quality: 1730.00              Length:    327
              Ratio:  5.307                Gaps:      1
 Percent Similarity:  99.694   Percent Identity: 99.694

alignment_block:
 US-09-357-675C-21 x US-60-278-258-200/rev   ..

 Align seg 1/1  to reverse of: US-60-278-258-200  from: 1  to: 3130

      2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   3034 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 2985

     18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2984 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCA 2935

     35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2934 GAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTG 2885

     52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2884 TGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGC 2835

     68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2834 TGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 2785

     85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2784 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 2735

    102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2734 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAG 2685

    118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2684 GGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAG 2635

    135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   2634 ACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGC 2585
```

```
     152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2584 AAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 2535

     168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2534 GATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 2485

     185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2484 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 2435

     202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2434 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGC 2385

     218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2384 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCC 2335

     235 roAlaHisTrpGluValLeuLeuArgAla.ArgAlaIleGluThrGlnCy 251
         |||||||||||||||||||||||||||| ||||||||||||||||||||
    2334 CAGCCCACTGGGAGGTGTTGCTGCGGGCCCCGTGCTATCGAAACCCAGTG 2285

     251 sTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaS 268
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2284 CTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAA 2235

     268 erTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArg 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2234 GTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGC 2185

     285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLe 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2184 TGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCT 2135

     301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2134 GCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACC 2085

     318 euTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||
    2084 TCTATGGCAATCTGGGTCACCCACTGTCT 2056

seq_name: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:PCT-US00-05988-692

seq_documentation_block:
; Sequence 692, Application PC/TUS0005988
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT: Steve Ruben
;  TITLE OF INVENTION: Human Prostate Cancer Associated Gene Sequences and Polypeptides
;  FILE REFERENCE: PA101PCT
;  CURRENT APPLICATION NUMBER: PCT/US00/05988
;  CURRENT FILING DATE: 2000-03-08
;  EARLIER APPLICATION NUMBER: 60/124,270
;  EARLIER FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
;   LENGTH: 1382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
PCT-US00-05988-692


alignment_scores:
            Quality: 1718.00            Length:    327
              Ratio:  5.270              Gaps:      0
 Percent Similarity:  99.694   Percent Identity: 99.694

alignment_block:
 US-09-357-675C-21 x PCT-US00-05988-692   ..

Align seg 1/1  to: PCT-US00-05988-692  from: 1  to: 1382

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 126

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 176

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     177 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 226

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     227 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 276

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     277 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 326

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     327 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

     101 LeuSerGluProLeuGlyGlyLysLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 426

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     427 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 476

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     477 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 526

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     527 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 576

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         |||||||||||||||||||  ||||||||||||||||||||||||||||
     577 AGAGATTCCAGGGCAGGGGCT.ATGTGTGAAAGCAACTCTACCATGCCTG 625

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     626 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 675

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     676 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 725

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     726 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 775

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     776 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 825

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     826 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 875

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     876 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 925

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

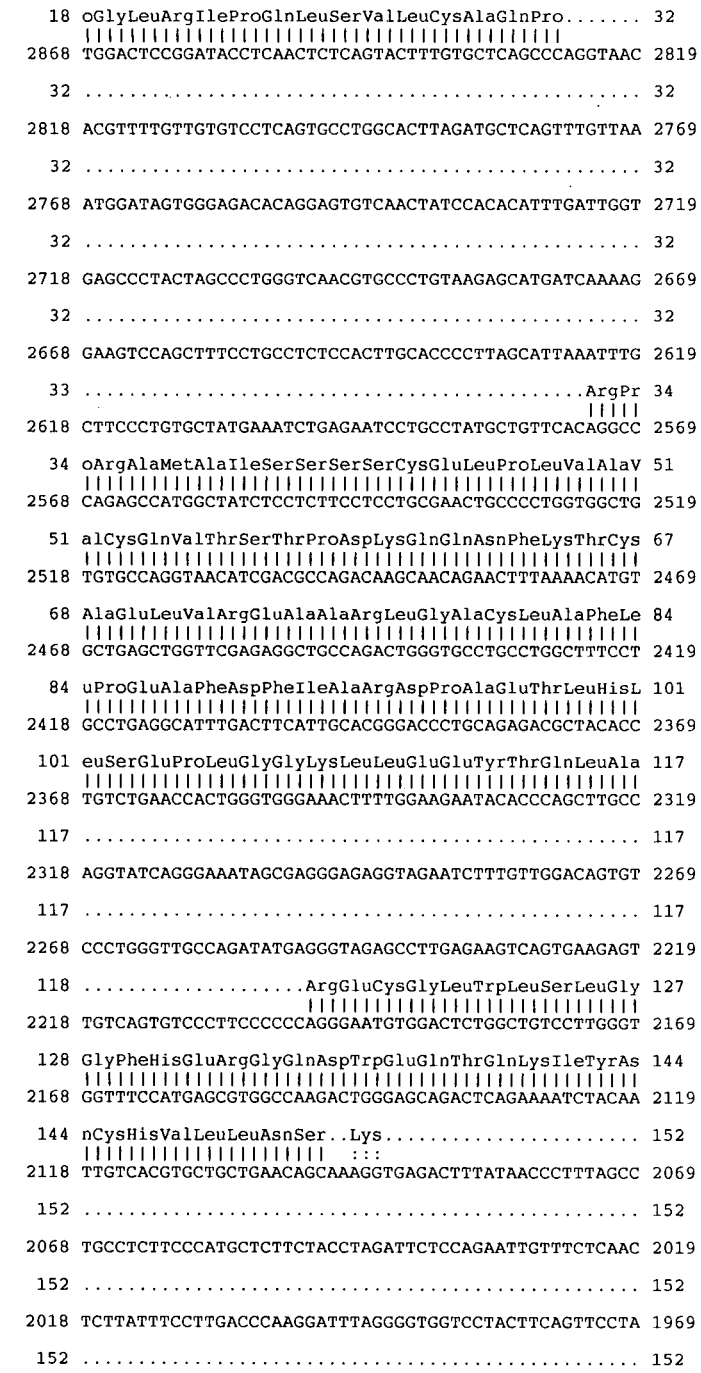
```
     926 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 975

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     976 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1025

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||
    1026 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1056

seq_name: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:US-09-925-300-692

seq_documentation_block:
; Sequence 692, Application US/09925300
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
;   LENGTH: 1382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-300-692


alignment_scores:
           Quality: 1718.00            Length:    327
             Ratio:  5.270              Gaps:      0
 Percent Similarity:  99.694   Percent Identity: 99.694

alignment_block:
 US-09-357-675C-21 x US-09-925-300-692   ..

 Align seg 1/1  to: US-09-925-300-692  from: 1  to: 1382

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 126

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 176

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     177 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 226

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     227 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 276

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     277 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 326

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     327 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 426

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     427 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 476

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     477 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 526

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     527 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 576

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         |||||||||||||||||||   ||||||||||||||||||||||||||||
     577 AGAGATTCCAGGGCAGGGGCT.ATGTGTGAAAGCAACTCTACCATGCCTG 625

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     626 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 675

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     676 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 725

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     726 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 775

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     776 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 825

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     826 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 875

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||| ||||||||||||||||||||||||||||||||||||||
     876 AAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 925

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     926 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 975

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     976 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1025

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||
    1026 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1056

seq_name: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:US-60-172-373-14716

seq_documentation_block:
; Sequence 14716, Application US/60172373
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT: Lal, Preeti
;  APPLICANT: Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym
;  FILE REFERENCE: GX-0006 P
;  CURRENT APPLICATION NUMBER: US/60/172,373
;  CURRENT FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 25,772
;  SOFTWARE: PERL Program
; SEQ ID NO 14716
;   LENGTH: 5035
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 480310.9c
```

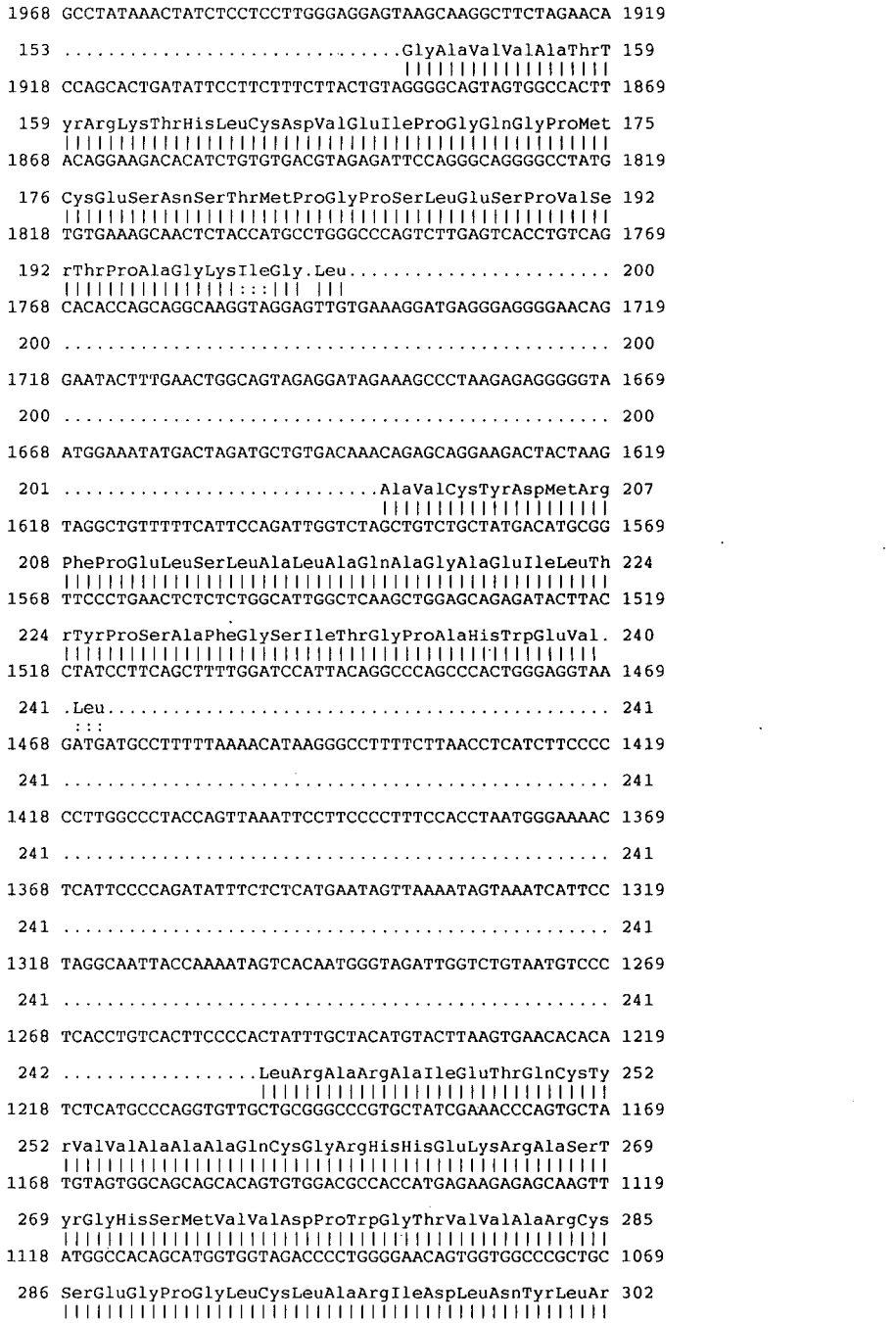
```
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 4170-4206
;   OTHER INFORMATION: a, t, c, g, or other
US-60-172-373-14716


alignment_scores:
            Quality: 1718.00              Length:    329
              Ratio:  5.254                Gaps:      2
 Percent Similarity:  99.392   Percent Identity: 99.088

alignment_block:
 US-09-357-675C-21 x US-60-172-373-14716   ..

 Align seg 1/1  to: US-60-172-373-14716  from: 1  to: 5035

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeu.LeuC 17
         |||||||||||||||||||||||||||||||||||||||||||| ||||
     669 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTTCTGT 718

      17 ysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArg 33
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     719 GTCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGG 768

      34 ProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAl 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     769 CCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGC 818

      50 aValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrC 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     819 TGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACAT 868

      67 ysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     869 GTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTC 918

      84 LeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHi 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     919 CTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACA 968

     100 sLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     969 CCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTG 1018

     117 laArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1019 CCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGC 1068

     134 GlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1069 CAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAA 1118

     150 nSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAsp. 166
         ||||||||||||||||||||||||||||||||||||||||||||||:::
    1119 CAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGAAC 1168

     167 ValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1169 GTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCC 1218

     183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1219 TGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTC 1268

     200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1269 TAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCT 1318

     217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleTh 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1319 CAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTAC 1368

     233 rGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1369 AGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCC 1418

     250 lnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArg 266
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1419 AGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGA 1468

     267 AlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAl 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1469 GCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGC 1518

     283 aArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnT 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1519 CCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACT 1568

     300 yrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgPro 316
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1569 ATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCT 1618

     317 AspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||||
    1619 GACCTCTATGGCAATCTGGGTCACCCACTGTCT 1651

seq_name: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:US-60-164-285-3819

seq_documentation_block:
; Sequence 3819, Application US/60164285
; GENERAL INFORMATION:
;  APPLICANT: Ma, Xiao-Jun
;  TITLE OF INVENTION: Tumor Associated Molecules (TAMs): Targets for diagnosis, trea
;  FILE REFERENCE: 3214
;  CURRENT APPLICATION NUMBER: US/60/164,285
;  CURRENT FILING DATE: 1999-11-05
;  NUMBER OF SEQ ID NOS: 8259
; SEQ ID NO 3819
;   LENGTH: 3917
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-164-285-3819


alignment_scores:
            Quality: 1717.00              Length:    328
              Ratio:  5.267                Gaps:      2
 Percent Similarity:  99.390   Percent Identity: 99.390

alignment_block:
 US-09-357-675C-21 x US-60-164-285-3819   ..

 Align seg 1/1  to: US-60-164-285-3819  from: 1  to: 3917

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     375 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 424

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     425 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCA 474

      35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     475 GAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTG 524

      52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     525 TGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGC 574

      68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     575 TGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGC 624

      85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 625 CTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 674

 102 SerGluProLeuGlyGlyLysLeuGluGluTyrThrGlnLeuAlaAr 118
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 675 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAG 724

 118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 725 GGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAG 774

 135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 775 ACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGC 824

 152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 825 AAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 874

 168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 875 GATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTGGGC 924

 185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 925 CCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 974

 202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 975 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGC 1024

 218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1025 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCC 1074

 235 roAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnCy 251
     |||||||||| ||||||||||||||||||||||||||||||||||||||||
1075 CAGCCCACTGGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTG 1124

 251 sTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaS 268
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1125 CTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAA 1174

 268 erTyrGlyHisSerMetValValAspProTrp.GlyThrValValAlaAr 284
     |||||||||||||||||||||||||||||| ||||||||||||||||||
1175 GTTATGGCCACAGCATGGTGGTAGACCCCTGGGGGAACAGTGGTGGCCCG 1224

 284 gCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1225 CTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATC 1274

 301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1275 TGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGAC 1324

 318 LeuTyrGlyAsnLeuGlyHisProLeuSer 327
     |||||||||||||||||||||||||||||
1325 CTCTATGGCAATCTGGGTCACCCACTGTCT 1354

seq_name: /cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-359-922-12822

seq_documentation_block:
; Sequence 12822, Application US/09359922
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cDNA
;  TITLE OF INVENTION: LIBRARIES
;  FILE REFERENCE: 20411-752CON1
;  CURRENT APPLICATION NUMBER: US/09/359,922
;  CURRENT FILING DATE: 1999-07-22
;  EARLIER APPLICATION NUMBER: US 09/205,155
;  EARLIER FILING DATE: 1998-12-03
;  NUMBER OF SEQ ID NOS: 13203

;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12822
;   LENGTH: 3724
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-359-922-12822


alignment_scores:
           Quality: 1626.50            Length:    331
             Ratio:  5.036              Gaps:      7
 Percent Similarity:  97.583   Percent Identity: 96.677

alignment_block:
 US-09-357-675C-21 x US-09-359-922-12822   ..

 Align seg 1/1  to: US-09-359-922-12822  from: 1  to: 3724

       4 PheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     383 TTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACT 432

      20 uArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaM 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     433 CCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCA 482

      37 etAlaIleSerSerSerSerCysGluLeuProLeu.ValAlaValCysGl 53
         |||||||||||||||||||||||||||||||| ||||||||||||||||
     483 TGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGGTGGCTGTGTGCCA 532

      53 nValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluL 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     533 GGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGCTGAGC 582

      70 euValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlu 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     583 TGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAG 632

      87 AlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGl 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     633 GCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGA 682

     103 uProLeu...GlyGlyLysLeuLeuGlu.GluTyrThrGlnLeuAlaArg 118
         |||||||   ||||||::::::   ||| ||||||||||||||||||||
     683 ACCACTGGGTGGGGGAAACTTTTGGGAAGGAATACACCCAGCTTGCCAGG 732

     119 GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAs 135
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     733 GAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGA 782

     135 pTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerL 152
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     783 CTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCA 832

     152 ysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGlu 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     833 AAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAG 882

     169 IlePro.GlyGlnGlyPro.MetCysGluSerAsnSerThrMetProGly 184
         |||||| |||||||||||| :::||||||||||||||||||||||||||||
     883 ATTCCAGGGGCAGGGGCCTATTGTGTGAAAGCAACTCTACCATGCCTGGG 932

     185 ProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     933 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 982

     201 aValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnA 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     983 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAG 1032

     218 laGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGly 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1033 CTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGC 1082
```

```
     235 ProAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnC 251
         |||||||||||| ||||||||||||||||||||||||||||||||||||||
    1083 CCAGCCCACTGGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGT 1132

     251 ysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAla 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1133 GCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCA 1182

     268 SerTyr.GlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         |||||| |||||||||||||||||||||||||||||||||||||||||||
    1183 AGTTATGGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 1232

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1233 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 1282

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1283 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1332

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1333 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1363

seq_name: /cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-359-922-12822

seq_documentation_block:
; Sequence 12822, Application US/09359922A
; GENERAL INFORMATION:
;  APPLICANT: Leshkowitz, Dena
;  APPLICANT: Liu, Jin
;  TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cDNA
;  TITLE OF INVENTION: LIBRARIES
;  FILE REFERENCE: 20411-752CON1
;  CURRENT APPLICATION NUMBER: US/09/359,922A
;  CURRENT FILING DATE: 1999-07-22
;  EARLIER APPLICATION NUMBER: US 09/205,155
;  EARLIER FILING DATE: 1998-12-03
;  EARLIER APPLICATION NUMBER: US 09/034,341
;  EARLIER FILING DATE: 1998-02-13
;  NUMBER OF SEQ ID NOS: 13203
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12822
;   LENGTH: 3724
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-359-922-12822


alignment_scores:
            Quality: 1626.50              Length:    331
              Ratio:  5.036                Gaps:      7
 Percent Similarity:  97.583   Percent Identity: 96.677

alignment_block:
 US-09-357-675C-21 x US-09-359-922-12822   ..

 Align seg 1/1  to: US-09-359-922-12822  from: 1  to: 3724

       4 PheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLe 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     383 TTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACT 432

      20 uArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaM 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     433 CCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCA 482

      37 etAlaIleSerSerSerSerCysGluLeuProLeu.ValAlaValCysGl 53
         |||||||||||||||||||||||||||||||||| ||||||||||||||
     483 TGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGGTGGCTGTGTGCCA 532

      53 nValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluL 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     533 GGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGTGCTGAGC 582

      70 euValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlu 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     583 TGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAG 632

      87 AlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGl 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     633 GCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGA 682

     103 uProLeu...GlyGlyLysLeuLeuGlu.GluTyrThrGlnLeuAlaArg 118
         |||||||   ||||||::::::   ||| ||||||||||||||||||||
     683 ACCACTGGGTGGGGGAAACTTTTGGGAAGGAATACACCCAGCTTGCCAGG 732

     119 GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAs 135
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     733 GAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGA 782

     135 pTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerL 152
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     783 CTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAACAGCA 832

     152 ysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGlu 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     833 AAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAG 882

     169 IlePro.GlyGlnGlyPro.MetCysGluSerAsnSerThrMetProGly 184
         |||||| |||||||||||| :::|||||||||||||||||||||||||||||
     883 ATTCCAGGGGCAGGGGCCTATTGTGTGAAAGCAACTCTACCATGCCTGGG 932

     185 ProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     933 CCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGC 982

     201 aValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnA 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     983 TGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAG 1032

     218 laGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGly 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1033 CTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGC 1082

     235 ProAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnC 251
         |||||||||||| ||||||||||||||||||||||||||||||||||||||
    1083 CCAGCCCACTGGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGT 1132

     251 ysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAla 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1133 GCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCA 1182

     268 SerTyr.GlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         |||||| |||||||||||||||||||||||||||||||||||||||||||
    1183 AGTTATGGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCC 1232

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1233 GCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 1282

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1283 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1332

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||
    1333 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1363

seq_name: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:US-60-360-207-3686

seq_documentation_block:
; Sequence 3686, Application US/60360207
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig
```

```
;  TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
;  FILE REFERENCE: CL001321
;  CURRENT APPLICATION NUMBER: US/60/360,207
;  CURRENT FILING DATE:  2002-03-01
;  NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 3686
;   LENGTH: 1296
;   TYPE: DNA
;   ORGANISM: HUMAN
US-60-360-207-3686


alignment_scores:
            Quality: 1487.00              Length:    327
              Ratio:  4.828                Gaps:      2
 Percent Similarity:  94.190   Percent Identity: 84.404

alignment_block:
 US-09-357-675C-21 x US-60-360-207-3686   ..

 Align seg 1/1  to: US-60-360-207-3686  from: 1  to: 1296

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||:::         ||||||||
      58 ATGCTGGGCTTCATCACCAGGCCTCCTCACCAA.........CTCCTGTG 98

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         |   |||   |||:::   ::::::   |||||||||:::||||||||||
      99 TACCGGATACCGATTACTTCGAATCCCAGTACTTTGTACTCAGCCCAGGC 148

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||:::|||:::   ||||||:::|||   ||||||||||||||||||
     149 CCAGAACCATGTCC...TCATCAACTTCCTGGGAGCTGCCCCTGGTGGCT 195

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||:::||||||:::|||||||||||||||
     196 GTGTGCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACATG 245

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
     246 TGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCCTTTC 295

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         |||||||||||||||||||||||||||:::|||||||||||||||||
     296 TGCCTGAGGCATTTGACTTTATTGCACGAAACCCTGCCGAGACATTACTC 345

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||:::|||   ||||||   :::|||:::||||||||
     346 CTGTCCGAACCACTGAATGGGGATCTTTTGGGCCAATATAGCCAGCTTGC 395

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||:::||||||||||||||||||||||||||||||||||
     396 CAGGGAATGTGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCC 445

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
     446 AAGACTGGGAGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAAC 495

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         |||||||||:::|||||||||:::|||||||||||||||||||||||||
     496 AGCAAGGGATCAGTAGTGGCCAGTTACAGGAAGACACATCTGTGCGATGT 545

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         |||||||||||||||||||||||||   |||||||||   |||   ||||
     546 AGAGATCCCAGGTCAGGGGCCGATGAGAGAAAGCAACTATACCAAGCCTG 595

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||   :::|||||||   |||||||:::|||||||||||||:::|||||||
     596 GAGGCACTCTTGAGCCACCTGTCAAGACACCGGCTGGCAAGGTTGGTCTA 645

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         |||:::|||||||||||||||||||||||||||||||   ||||||||
     646 GCAATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTCA 695

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||:::||||
     696 AGCTGGGGCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTACAG 745

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
     746 GTCCAGCCCACTGGGAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAG 795

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         |||||||||:::|||||||||||||||||||||||||||||   |||||
     796 TGCTATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGC 845

     267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     846 AAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCC 895

     284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
         ||||||||||||||||||||||||||||||||||||||||||||::::::
     896 GCTGCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTT 945

     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
         |||:::|||:::|||:::||||||||||||||||||||||||||||||||
     946 CTACAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGA 995

     317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         ||||||||||:::|||||||||||||||||||
     996 CCTCTATGGCAGTCTGGGTCATCCACTCTCT 1026

seq_name: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:US-60-242-679-78

seq_documentation_block:
; Sequence 78, Application US/60242679
; GENERAL INFORMATION:
;  APPLICANT:  Ladunga, Steven Istvan
;  APPLICANT:  Spier, Eugene
;  APPLICANT:  Greenberg, Simon
;  APPLICANT:  Brandenberger, Ralph
;  APPLICANT:  Wang, Yu
;  APPLICANT:  Dubman, Alex
;  TITLE OF INVENTION:  ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL000898-PROV
;  CURRENT APPLICATION NUMBER: US/60/242,679
;  CURRENT FILING DATE:  2000-10-24
;  NUMBER OF SEQ ID NOS:  2265
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 78
;   LENGTH: 3620
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3620)
;   OTHER INFORMATION: n = A,T,C or G
US-60-242-679-78


alignment_scores:
            Quality: 1468.50              Length:    660
              Ratio:  4.505                Gaps:      8
 Percent Similarity:  49.394   Percent Identity: 48.939

alignment_block:
 US-09-357-675C-21 x US-60-242-679-78/rev   ..

 Align seg 1/1  to reverse of: US-60-242-679-78  from: 1  to: 3620

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2918 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCC 2869
```


18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
2868 TGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGTAAC 2819
32 32
2818 ACGTTTTGTTGTGTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTTAA 2769
32 32
2768 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 2719
32 32
2718 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 2669
32 32
2668 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 2619
33 .....ArgPr 34
2618 CTTCCCTGTGCTATGAAATCTGAGAATCCTGCCTATGCTGTTCACAGGCC 2569
34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
2568 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 2519
51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
2518 TGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 2469
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
2468 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCT 2419
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
2418 GCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACC 2369
101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
2368 TGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCC 2319
117 117
2318 AGGTATCAGGGAAATAGCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 2269
117 117
2268 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 2219
118 .....ArgGluCysGlyLeuTrpLeuSerLeuGly 127
2218 TGTCAGTGTCCCTTCCCCCCAGGGAATGTGGACTCTGGCTGTCCTTGGGT 2169
128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
2168 GGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAA 2119
144 nCysHisValLeuLeuAsnSer..Lys..... 152
2118 TTGTCACGTGCTGCTGAACAGCAAAGGTGAGACTTTATAACCCTTTAGCC 2069
152 152
2068 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAATTGTTTCTCAAC 2019
152 152
2018 TCTTATTTCCTTGACCCAAGGATTTAGGGGTGGTCCTACTTCAGTTCCTA 1969
152 152


1968 GCCTATAAACTATCTCCTCCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 1919
153 .....GlyAlaValValAlaThrT 159
1918 CCAGCACTGATATTCCTTCTTTCTTACTGTAGGGGCAGTAGTGGCCACTT 1869
159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
1868 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 1819
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
1818 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 1769
192 rThrProAlaGlyLysIleGly.Leu..... 200
1768 CACACCAGCAGGCAAGGTAGGAGTTGTGAAAGGATGAGGGAGGGGAACAG 1719
200 200
1718 GAATACTTTGAACTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGGTA 1669
200 200
1668 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 1619
201 .....AlaValCysTyrAspMetArg 207
1618 TAGGCTGTTTTTCATTCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 1569
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
1568 TTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 1519
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
1518 CTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTAA 1469
241 .Leu..... 241
1468 GATGATGCCTTTTTAAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 1419
241 241
1418 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAAC 1369
241 241
1368 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTCC 1319
241 241
1318 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 1269
241 241
1268 TCACCTGTCACTTCCCCACTATTTGCTACATGTACTTAAGTGAACACACA 1219
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTy 252
1218 TCTCATGCCCAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTA 1169
252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
1168 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 1119
269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
1118 ATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 1069
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302

```
    1068 TCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTGCG 1019

     302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1018 ACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCT 969

     319 yrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||
     968 ATGGCAATCTGGGTCACCCACTGTCT 943

seq_name: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:US-60-324-185-10674

seq_documentation_block:
; Sequence 10674, Application US/60324185
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
;  TITLE OF INVENTION:  POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
;  TITLE OF INVENTION:  POLYMORPHISMS IDENTIFIED THEREBY
;  FILE REFERENCE: GX-0019-1 P
;  CURRENT APPLICATION NUMBER: US/60/324,185
;  CURRENT FILING DATE:  2001-09-21
;  NUMBER OF SEQ ID NOS: 35862
;  SOFTWARE: PERL Program
; SEQ ID NO 10674
;   LENGTH: 5482
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 1398463.12
;   NAME/KEY: unsure
;   LOCATION: 1652, 4633, 5437
;   OTHER INFORMATION: a, t, c, g, or other
US-60-324-185-10674


alignment_scores:
            Quality: 1315.00              Length:    354
              Ratio:  4.242                Gaps:     25
 Percent Similarity:  87.571   Percent Identity: 86.441

alignment_block:
 US-09-357-675C-21 x US-60-324-185-10674   ..

 Align seg 1/1  to: US-60-324-185-10674  from: 1  to: 5482

       2 LeuGlyPheIleThrArg.ProProHisArgPheLeuSerLeuLeuCysP 18
         |||||||||||||||||    |||||||||||||||||||||||||||||
     680 CTGGGCTTCATCACCAGGGGCTCCTCACAGATTCCTGTCCCTTCTGTGTC 729

      18 roGlyLeuArgIleProGlnLeuSerValLeuCys.AlaGlnProArgPr 34
         |||||||||||||||||||||||||||||||||| ||||||||||||||
     730 CTGGACTCCGGATACCTCAACTCTCAGTACTTTGTTGCTCAGCCCAGGCC 779

      34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaV 51
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     780 CAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCTG 829

      51 alCysGln.ValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         |||||||| |||||||||||||||||||||||||||||||||||||||||
     830 TGTGCCAGGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 879

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     880 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 929

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     930 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 979

     101 LeuSerGluProLeuGly.GlyLysLeuLeuGluGluTyrThrGlnLeuA 117
         |||||||||||||||||| ||||||||||||||||||||||||||||||||
     980 CTGTCTGAACCACTGGGTTGGGAAACTTTTGGAAGAATACACCCAGCTTG 1029

     117 laArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArg.Gl 133
         |||||||||||||||||||||||||||||||||||||||||||||||| ||
    1030 CCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGG 1079

     133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1080 CCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGA 1129

     150 snSerLysGlyAlaValVal.AlaThr...TyrArgLysThrHisLeuCy 165
         |||||||||||||||||||| ||||||         :::|||||||||||
    1130 ACAGCAAAGGGGCAGTAGTGGGCCACTTAACAGGGAAGGACACATCTGTG 1179

     165 sAsp.ValGluIlePro.GlyGlnGlyProMetCysGluSerAsnSerTh 181
         |::: |||||||||||| ||||||||||||||||||||||||||||||||
    1180 TGAACGTAGAGATTCCAAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 1229

     181 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1230 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGA 1279

     198 leGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 214
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1280 TTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCA 1329

     215 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySe 231
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1330 TTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATC 1379

     231 rIleThrGlyProAlaHis.TrpGlu.ValLeuLeu.ArgAlaArgAlaI 247
         |||||||||||||||||| |||||| |||||||| ||||||
    1380 CATTACAGGCCCAGCCCACTTGGGAAGGTGTTGCTGGCGGGCC..... 1422

     247 leGluThrGlnCysTyr.....................ValValAlaAla 256
                       ||||||                      :::||||||
    1423 ........CCGTGCTATCCGAAACCCAGTCGCTTATGTAAGTTGGCAGCA 1464

     257 AlaGlnCys.GlyArg.HisHisGluLysArgAlaSerTyrGlyHisSer 272
         ||||||||| |||||| ||||||||||||||||||||||||||||||||||
    1465 GCACAGTGTGGGACGGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGC 1514

     273 Met.ValValAspProTrp.GlyThrValValAlaArgCysSerGlu.Gl 288
         ||| |||||||||||||| |||||||||||||||||||||||     ||
    1515 ATGGGTGGTAGACCCCTGGGGGAACAGTGGTGGCCCGCTGCTCTGTAGGG 1564

     288 yProGlyLeuCys.LeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeu 304
         |||||||||||| ||||||||||||||||||||||||||||||||:::
    1565 GCCAGGCCTCTGCCCTTGCCCGAATAGACCTCAACTATCTGCGACAAGTT 1614

     305 ArgArgHis...LeuProVal.PheGlnHisArgArg.ProAsp.LeuTy 319
                 |||||||||| |||||||||||||| |||||| |||||
    1615 GCGCCTGACACCCTGCCTGTTGTTCCAGCACCGCAGGNCCTGACCCTCTA 1664

     319 rGlyAsn.LeuGlyHisPro 325
         |||||||| ||||||||||||
    1665 TGGCAATCCTGGGTCACCCC 1684
```

```
OM of: US-09-357-675C-21 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Apr 29, 2002 12:20 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091758_3352/app_query.fasta_1.1208
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=p2n.rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09357675_@CGN1_1_88 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: Pending_Patents_NA_New:*
Database sequences: 139564
Database length: 91993600
Search time (sec): 106.130000



score_list:
Sequence                                    Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:US-60-365-384-491 -  491.50  705.56  4.8e-32   1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-235 +  259.50  370.09  2.3e-13
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-33512 +  233.00  341.77  8.8e-12
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-19553 +  189.50  280.31  2.3e-08
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-466 +  185.00  255.76  5.4e-07
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-102-524-748 +  129.50  191.56  0.0020    2
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-1148 +  126.50  177.18  0.0129
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-5721 +  124.00  184.19  0.0053
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-5426 +  118.50  176.78  0.0136
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-975-254-11297 +  107.00  158.93  0.1340
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-115-123-90 +   95.00  128.46  6.67     10
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-14411 +   93.00  102.16  194.66
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-14786 +   93.00  102.16  194.66
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-108 +   92.50  121.10  17.15     1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-1860 +   92.50  118.13  25.10
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-103-295-384 -   92.50   96.83  385.55    1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-103-295-385 -   92.50   96.61  396.24    2
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-35282 +   91.50  135.12  2.84
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:US-60-365-384-61 +   91.50  118.80  23.01     16
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-109-528-1 +   91.00  110.19  69.46     385
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-115-123-126 +   89.50  118.53  23.84     1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-108-605-74 +   89.50  114.43  40.34     19
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-413-186A-3 +   89.50  108.23  89.38     37
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-10831 +   89.00  103.71  159.51
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:US-60-365-384-472 +   88.50  100.88  229.27    7
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-627-465B-2 +   87.50  113.52  45.31     15
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-118-328-1 +   87.50  110.69  65.16     213
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-14118 -   87.50   82.18  2.5e+03
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-975-254-8983 +   87.00  130.91  4.87
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-102-806-80 -   87.00  106.11  117.20    32
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-108-605-210 -   87.00  100.22  249.51    5
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-108-605-212 -   87.00   99.74  265.42    6
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-8734 +   87.00   99.00  291.95
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-9557 +   86.50  104.25  148.81
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-7442 -   86.50   99.98  257.33
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-9491 +   86.50   97.67  346.14
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-107-431-280 -   86.50   79.72  3.4e+03    4
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-6571 -   86.00  113.91  43.11
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-12023 +   86.00  100.11  253.10
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-12021 +   86.00   97.75  342.54
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-12019 +   86.00   97.11  371.91
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-11976 -   86.00   91.99  716.98
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-15008 -   86.00   90.03  922.16
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-15009 -   86.00   82.32  2.5e+0
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-11975 -   86.00   78.17  4.2e+0

seq_name: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:US-60-365-384-491

seq_documentation_block:
; Sequence 491, Application US/60365384
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Wang, Jian-rui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Ghosh, Malabika
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Weng, Gezhi
;  TITLE OF INVENTION: Novel Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 814
;  CURRENT APPLICATION NUMBER: US/60/365,384
;  CURRENT FILING DATE:  2002-03-14
;  NUMBER OF SEQ ID NOS: 666
;  SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 491
;   LENGTH: 1251
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-365-384-491


alignment_scores:
           Quality:  491.50              Length:    312
             Ratio:  2.398                Gaps:     10
 Percent Similarity:  65.705   Percent Identity: 36.218

alignment_block:
 US-09-357-675C-21 x US-60-365-384-491/rev   ..

 Align seg 1/1  to reverse of: US-60-365-384-491  from: 1  to: 1251

      23 ProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIl 39
         ||||||:::   :::::::|||               |||:::|||
    1201 CCGCAGGTGGTGCTTGTCTGC..............AGAGTCATG..... 1172

      39 eSerSerSerSerCysGluLeuProLeuValAlaValCysGlnValThrS 56
              :::|||   :::|||   |||:::   :::   |||::::::|
    1171 .......ACCTCTTTCCGCTTGGCCCTCATCCAGCTT...CAGATTTCTT 1132

      56 erThrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArg 72
         ||         |||:::::::|||         ::::::::::::::|||
    1131 CCATC......AAATCAGATAACGTCACTCGCGCTTGTAGCTTCATCCGG 1088

      73 GluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAs 89
         |||||||||      ||||||   ::::::   |||||||||:::|||
    1087 GAGGCAGCAACGCAAGGAGCCAAAATAGTTTCTTTGCCGGAATGCTTT.. 1040

      89 pPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 106
                                           :::::::|||   |
    1039 .....................................AATTCTCCATATG 1027

     106 lyGlyLysLeuGluGluTyr..................... 113
         ||:::|||   :::   ||||||
    1026 GAGCGAAATATTTTCCTGAATATGCAGAGAAAATTCCTGGTGAATCCACA 977
```

```
     114 .........ThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
                 :::::::::|||:::||||||::::::::::|||      ||
     976 CAGAAGCTTTCTGAAGTAGCAAAGGAATGCAGCATATATCTCATTGGAGG 927

     127 yGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrA 144
         |::::::   |||             |||::::::   |||:::||||
     926 CTCTATCCCTGAA..............GAGGATGCTGGGAAATTATATA 892

     144 snCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
         ||      ::::::::::::      |||:::::::::|||   ||||||
     891 ACACCTGTGCTGTGTTTGGGCCTGATGGAACTTTACTAGCAAAGTATAGA 842

     161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGl 177
         |||   ||||||   |||:::::::::||||||:::      :::   ||
     841 AAGATCCATCTGTTTGACATTGATGTTCCTGGAAAAATTACATTTCAAGA 792

     177 uSerAsnSerThrMetProGlyProSerLeuGluSerProValSerThrP 194
         ||||::::::      ||||||   |||:::   |||      :::||||
     791 ATCTAAAACATTGAGTCCGGGTGATAGTTTC...TCCACATTTGATACTC 745

     194 roAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGlu 210
         ||      ::::::||||||::::::|||||||||||||||||   |||
     744 CTTACTGCAGAGTGGGTCTGGGCATCTGCTACGACATGCGGTTTGCAGAG 695

     211 LeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSe 227
         |||:::         ||||||   |||:::::::::|||:::||||||::
     694 CTTGCACAAATCTACGCACAGAGAGGCTGCCAGCTGTTGGTATATCCAGG 645

     227 rAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgA 244
         :||||||:::      ||||||||||||||||||||||:::|||   |||:
     644 AGCTTTTAATCTGACCACTGGACCAGCCCATTGGGAGTTACTTCAGCGAA 595

     244 laArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGly 260
         ::||||||:::::::::|||   ||||||::::::|||:::   :::
     594 GCCGGGCTGTTGATAATCAGGTGTATGTGGCCACAGCCTCTCCTGCC... 548

     261 ArgHisHisGluLysArgAlaSerTyr.........GlyHisSerMetVa 274
               ::::::   :::|||||||||         |||||||||   ||
     547 ......CGGGATGACAAAGCCTCCTATGTTGCCTGGGGACACAGCACCGT 504

     274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
         ||||:::|||||||||   |||:::|||:::::::::      ::::
     503 GGTGAACCCTTGGGGGGAGGTTCTAGCCAAAGCTGGCACAGAAGAAGCAA 454

     291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
         ::      :::   ||||||||::   |||   ::::::|||::::::
     453 TCGTGTATTCAGACATAGACCTGAAGAAGCTGGCTGAAATACGCCAGCAA 404

     308 LeuProValPheGlnHisArgArgProAspLeuTyr 319
         :::|||||||||:::::::::|||   |||||||||
     403 ATCCCCGTTTTTAGACAGAAGCGATCAGACCTCTAT 368
```

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-235

```
seq_documentation_block:
; Sequence 235, Application US/09540209B
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1001-001
;  CURRENT APPLICATION NUMBER: US/09/540,209B
;  CURRENT FILING DATE:  2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 235
;   LENGTH: 888
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-09-540-209B-235
```

alignment_scores:

```
                Quality:  259.50          Length:    298
                  Ratio:  1.612             Gaps:      8
  Percent Similarity:  54.027   Percent Identity: 28.188

alignment_block:
 US-09-357-675C-21 x US-09-540-209B-235   ..

 Align seg 1/1  to: US-09-540-209B-235  from: 1  to: 888

      49 ValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLy 65
         |||::::::   |||   :::::::|||   |||   :::   |||:::
      19 GTAGGAATCATTCAACAGGCTAACACATCAGATATTAGGATAAACCTGAT 68

      65 sThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuA 82
          :::   |||:::   ::::::   :::|||      ||||||   |||:
      69 GAACCTGGCTAAAAGTATTGAAGCATGTGCCGCTAATGGCGCTCACCTTG 118

      82 laPheLeuProGluAlaPheAsp.........PheIleAlaArgAspPro 95
         ::   |||   |||      :::          |||         :::
     119 TTGTTCTGCAAGAACTTCATAATTCTTTGTATTTCTGTCAGACAGAGAAT 168

      96 AlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGl 112
         ::::::   ::::::|||:::||||||:::   |||
     169 ACGGATTTATTTGAACTGGCAGAACCCATTCCTGGCCCTTCTACCGGATT 218

     112 uTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyP 129
          |||::::::||||||            :::   |||      :::   |
     219 CTATTCCGAACTGGCGGCAGCCAATCGGATAGTGCTTGTTACTTCTTTGT 268

     129 heHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCys 145
         ||::::::|||:::                          :::|||
     269 TTGAGAAACGTGCTCCGGGA..............CTATATCATAATACA 303

     146 HisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArgLysTh 162
            |||:::::::::      |||:::::::::::   |||||||||
     304 GCTGTTGTCTTTGACCGGGATGGAAGTATTGCCGGAAAATATCGTAAGAT 353

     162 rHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerA 179
          |||               ||||||   :::            |||:::
     354 GCAT..............ATTCCTGATGATCCGGCTTATTACGAGAAAT 388

     179 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 195
                 ||||||            :::|||::::::|||
     389 TCTATTTTACTCCGGGAGATATTGGCTTTGAACCGATTCAGACCTCTTTA 438

     196 GlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSe 212
         ||||||:::|||:::   ||||||:::|||:::   :::||||||   ::
     439 GGCAAGTTGGGTGTGTTGGTTTGCTGGGATCAATGGTATCCGGAAGCTGC 488

     212 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaP 229
         :      :::|||      ||||||||||||||   ||||||:::|||:
     489 TCGCCTGATGGCGTTGAAAGGAGCTGAGATTTTGATTTATCCTACTGCTA 538

     229 heGlySerIleThrGlyProAlaHis..................... 237
         ::|||      :::      :::
     539 TCGGTTGGGAGAGTACAGATACAGATGACGAAAAGAAACGTCAGCTCAAT 588

     238 ...TrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa 253
            |||   :::      ||||||:::|||:::   :::          ||
     589 GCTTGGATTATTTCTCAGCGTGCGCATGCGGTAGCCAATGGGCTTCCGGT 638

     253 lValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSer.... 268
         |:::::::::   :::   |||      ||||||         |||
     639 GATTTCAGTCAATCGTGTCGGT......CACGAACCTGATCCGTCAGGAC 682

     269 ...................TyrGlyHisSerMetValValAspProTrp 278
                             :::|||:::|||:::|||:::   |||
     683 AGACCAACGGGATATTATTTTGGGGAAATAGTTTTGTTGCCGGACCGCAG 732

     279 GlyThrValValAlaArgCys............SerGluGlyProGlyLe 291
         |||       :::|||::::::           ||||||:::
     733 GGTGAATACCTGGCTCAGGCGGGAAATGACCGCTCTGAAAAT........ 774
```

```
     291 uCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisL 308
         ::::::::::::||||||:::      ::::::::::||||||
     775 .ATGATTGTTGAGGTGGATCTTGAACGTTCGGAGAATGTGCGTCGTTGGT 823

     308 euProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeu 322
           |||   ::::::   ||||||   |||   ||||||||||||
     824 GGCCATTTCTTCGTGATCGGAGGATAGATGAATATGGGAATTTA 867

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-33512

seq_documentation_block:
; Sequence 33512, Application US/09540210B
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
;  FILE REFERENCE: PD-1037 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,210B
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/972,899
;  PRIOR FILING DATE: November 18, 1997
;  PRIOR APPLICATION NUMBER: 08/395,244
;  PRIOR FILING DATE: February 27, 1995
;  PRIOR APPLICATION NUMBER: 08/722,922
;  PRIOR FILING DATE: September 27, 1996
;  PRIOR APPLICATION NUMBER: 60/005,526
;  PRIOR FILING DATE: September 29, 1995
;  PRIOR APPLICATION NUMBER: 08/824,029
;  PRIOR FILING DATE: March 25, 1997
;  PRIOR APPLICATION NUMBER: 60/014,010
;  PRIOR FILING DATE: March 25, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/903,555
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/862,178
;  PRIOR FILING DATE: May 22, 1997
;  PRIOR APPLICATION NUMBER: 60/018,217
;  PRIOR FILING DATE: May 23, 1996
;  PRIOR APPLICATION NUMBER: 08/881,589
;  PRIOR FILING DATE: June 24, 1997
;  PRIOR APPLICATION NUMBER: 60/021,275
;  PRIOR FILING DATE: June 25, 1996
;  PRIOR APPLICATION NUMBER: 08/903,802
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/905,881
;  PRIOR FILING DATE: August 1, 1997
;  PRIOR APPLICATION NUMBER: 60/025,204
;  PRIOR FILING DATE: August 1, 1996
;  PRIOR APPLICATION NUMBER: 08/903,471
;  PRIOR FILING DATE: July 30, 1997
;  PRIOR APPLICATION NUMBER: 60/025,478
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/903,556
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/025,217
;  PRIOR FILING DATE: August 22, 1996
;  PRIOR APPLICATION NUMBER: 08/937,142
;  PRIOR FILING DATE: September 23, 1997
;  PRIOR APPLICATION NUMBER: 60/026,598
;  PRIOR FILING DATE: September 24, 1996
;  PRIOR APPLICATION NUMBER: 08/960,746
;  PRIOR FILING DATE: October 29, 1997
;  PRIOR APPLICATION NUMBER: 60/030,144
;  PRIOR FILING DATE: October 30, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/755,524
;  PRIOR FILING DATE: November 22, 1996
;  PRIOR APPLICATION NUMBER: 60/007,495
;  PRIOR FILING DATE: November 22, 1995
;  PRIOR APPLICATION NUMBER: 09/021,031
;  PRIOR FILING DATE: February 10, 1998
;  PRIOR APPLICATION NUMBER: 60/039,325
;  PRIOR FILING DATE: February 13, 1997
;  PRIOR APPLICATION NUMBER: 09/035,172
;  PRIOR FILING DATE: March 4, 1998
;  PRIOR APPLICATION NUMBER: 60/040,431
;  PRIOR FILING DATE: March 5, 1997
;  PRIOR APPLICATION NUMBER: 09/041,894
;  PRIOR FILING DATE: March 12, 1998
;  PRIOR APPLICATION NUMBER: 60/040,199
;  PRIOR FILING DATE: March 14, 1997
;  PRIOR APPLICATION NUMBER: 09/050,817
;  PRIOR FILING DATE: March 30, 1998
;  PRIOR APPLICATION NUMBER: 60/043,792
;  PRIOR FILING DATE: April 11, 1997
;  PRIOR APPLICATION NUMBER: 09/074,999
;  PRIOR FILING DATE: May 8, 1998
;  PRIOR APPLICATION NUMBER: 60/048,431
;  PRIOR FILING DATE: May 29, 1997
;  PRIOR APPLICATION NUMBER: 09/107,592
;  PRIOR FILING DATE: June 30, 1998
;  PRIOR APPLICATION NUMBER: 60/052,751
;  PRIOR FILING DATE: July 1, 1997
;  PRIOR APPLICATION NUMBER: 09/094,079
;  PRIOR FILING DATE: June 9, 1998
;  PRIOR APPLICATION NUMBER: 60/049,975
;  PRIOR FILING DATE: June 13, 1997
;  NUMBER OF SEQ ID NOS: 35654
;  SOFTWARE: PERL Program
; SEQ ID NO 33512
;   LENGTH: 297
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: rat00219258
US-09-540-210B-33512


alignment_scores:
            Quality:  233.00              Length:    102
              Ratio:  2.949                Gaps:      2
 Percent Similarity:  77.451   Percent Identity: 43.137

alignment_block:
 US-09-357-675C-21 x US-09-540-210B-33512   ..

 Align seg 1/1  to: US-09-540-210B-33512  from: 1  to: 297

     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHi 237
         :::::::|||:::|||||||:::||||||:::      ||||||||||||||
       1 CAGCTCTTGGTGTATCCTGGAGCTTTCAATATGACCACTGGACCAGCCCA 50

     237 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValV 254
         |||||||:::|||   ||||||||||||:::::::::|||   ||||||:
      51 CTGGGAGTTGCTTCAGCGAGCCCGAGCTGTTGATAATCAGGTGTATGTGG 100

     254 alAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyr... 269
         ::::::|||:::   :::          ::::::::::::|||||||||
     101 CTACAGCCTCTCCTGCC.........CGGGATGAGAAAGCCTCGTACGTG 141

     270 ......GlyHisSerMetValValAspProTrpGlyThrValValAlaAr 284
```

```
            |||||||||   ||||||||||||||||||   |||::::::::
    142 GCCTGGGGACACAGCACTGTGGTGGATCCTTGGGGGCAGGTCCTAACCAA 191

    284 gCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
        :::::::           :::      :::   |||||||||:::   |
    192 AGCTGGCACCGAGGAAACAATCCTGTACTCAGACATAGACCTGAAGAAGC 241

    301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
        ||   ::::::|||:::::::::|||:::::::::::::::|||   |||
    242 TGTCTGAAATTCGGCAGCAAATTCCCATTTTAAAACAGAAACGAGCAGAC 291

    318 LeuTyr 319
        ||||||
    292 CTCTAT 297

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-19553

seq_documentation_block:
; Sequence 19553, Application US/09540210B
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
;  FILE REFERENCE: PD-1037 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,210B
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/972,899
;  PRIOR FILING DATE: November 18, 1997
;  PRIOR APPLICATION NUMBER: 08/395,244
;  PRIOR FILING DATE: February 27, 1995
;  PRIOR APPLICATION NUMBER: 08/722,922
;  PRIOR FILING DATE: September 27, 1996
;  PRIOR APPLICATION NUMBER: 60/005,526
;  PRIOR FILING DATE: September 29, 1995
;  PRIOR APPLICATION NUMBER: 08/824,029
;  PRIOR FILING DATE: March 25, 1997
;  PRIOR APPLICATION NUMBER: 60/014,010
;  PRIOR FILING DATE: March 25, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/903,555
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/862,178
;  PRIOR FILING DATE: May 22, 1997
;  PRIOR APPLICATION NUMBER: 60/018,217
;  PRIOR FILING DATE: May 23, 1996
;  PRIOR APPLICATION NUMBER: 08/881,589
;  PRIOR FILING DATE: June 24, 1997
;  PRIOR APPLICATION NUMBER: 60/021,275
;  PRIOR FILING DATE: June 25, 1996
;  PRIOR APPLICATION NUMBER: 08/903,802
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/905,881
;  PRIOR FILING DATE: August 1, 1997
;  PRIOR APPLICATION NUMBER: 60/025,204
;  PRIOR FILING DATE: August 1, 1996
;  PRIOR APPLICATION NUMBER: 08/903,471
;  PRIOR FILING DATE: July 30, 1997
;  PRIOR APPLICATION NUMBER: 60/025,478
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/903,556
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/025,217
;  PRIOR FILING DATE: August 22, 1996
;  PRIOR APPLICATION NUMBER: 08/937,142
;  PRIOR FILING DATE: September 23, 1997
;  PRIOR APPLICATION NUMBER: 60/026,598
;  PRIOR FILING DATE: September 24, 1996
;  PRIOR APPLICATION NUMBER: 08/960,746
;  PRIOR FILING DATE: October 29, 1997
;  PRIOR APPLICATION NUMBER: 60/030,144
;  PRIOR FILING DATE: October 30, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/755,524
;  PRIOR FILING DATE: November 22, 1996
;  PRIOR APPLICATION NUMBER: 60/007,495
;  PRIOR FILING DATE: November 22, 1995
;  PRIOR APPLICATION NUMBER: 09/021,031
;  PRIOR FILING DATE: February 10, 1998
;  PRIOR APPLICATION NUMBER: 60/039,325
;  PRIOR FILING DATE: February 13, 1997
;  PRIOR APPLICATION NUMBER: 09/035,172
;  PRIOR FILING DATE: March 4, 1998
;  PRIOR APPLICATION NUMBER: 60/040,431
;  PRIOR FILING DATE: March 5, 1997
;  PRIOR APPLICATION NUMBER: 09/041,894
;  PRIOR FILING DATE: March 12, 1998
;  PRIOR APPLICATION NUMBER: 60/040,199
;  PRIOR FILING DATE: March 14, 1997
;  PRIOR APPLICATION NUMBER: 09/050,817
;  PRIOR FILING DATE: March 30, 1998
;  PRIOR APPLICATION NUMBER: 60/043,792
;  PRIOR FILING DATE: April 11, 1997
;  PRIOR APPLICATION NUMBER: 09/074,999
;  PRIOR FILING DATE: May 8, 1998
;  PRIOR APPLICATION NUMBER: 60/048,431
;  PRIOR FILING DATE: May 29, 1997
;  PRIOR APPLICATION NUMBER: 09/107,592
;  PRIOR FILING DATE: June 30, 1998
;  PRIOR APPLICATION NUMBER: 60/052,751
;  PRIOR FILING DATE: July 1, 1997
;  PRIOR APPLICATION NUMBER: 09/094,079
;  PRIOR FILING DATE: June 9, 1998
;  PRIOR APPLICATION NUMBER: 60/049,975
;  PRIOR FILING DATE: June 13, 1997
;  NUMBER OF SEQ ID NOS: 35654
;  SOFTWARE: PERL Program
; SEQ ID NO 19553
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00309797
;   NAME/KEY: unsure
;   LOCATION: 224
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-19553


alignment_scores:
            Quality:  189.50              Length:     75
              Ratio:  3.384                Gaps:      1
 Percent Similarity:  74.667   Percent Identity: 49.333

alignment_block:
 US-09-357-675C-21 x US-09-540-210B-19553   ..

 Align seg 1/1  to: US-09-540-210B-19553  from: 1  to: 239

    169 IleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyPr 185
        :::||||||:::      :::   ||||||::::::      ||||||
      1 GTTCCTGGAAAAATTACATTTCAAGAATCTAAAACATTGAGTCCGGGTGA 50
```

```
     185 oSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaV 202
          |||:::   |||      :::||||||      ::::::||||||::::
      51 TAGTTTC...TCCACATTTGATACTCCTTACTGCAGAGTGGGTCTGGGCA 97

     202 alCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAla 218
         ::||||||||||||||||||   ||||||:::          ||||||
      98 TCTGCTACGACATGCGGTTTGCAGAGCTTGCACAAATCTACGCACAGAGA 147

     219 GlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyPr 235
         |||:::::::::|||:::||||||:::|||||||:::      ||||||||
     148 GGCTGCCAGCTGTTGGTATATCCAGGAGCTTTTAATCTGACCACTGGACC 197

     235 oAlaHisTrpGluValLeuLeuArg 243
         ||||||||||||||:::|||   |||
     198 AGCCCATTGGGAGTTACTTCAGCGA 222

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-466

seq_documentation_block:
; Sequence 466, Application US/09540209B
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1001-001
;  CURRENT APPLICATION NUMBER: US/09/540,209B
;  CURRENT FILING DATE:  2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 466
;   LENGTH: 1599
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-09-540-209B-466


alignment_scores:
           Quality:  185.00            Length:    249
             Ratio:  1.350               Gaps:     11
 Percent Similarity:  55.020   Percent Identity: 28.112

alignment_block:
 US-09-357-675C-21 x US-09-540-209B-466   ..

 Align seg 1/1  to: US-09-540-209B-466  from: 1  to: 1599   .

      85 ProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLe 101
         |||:::|||      :::      |||:::|||   ::::::||||||
     918 CCGCGGGCTGGCCCAATATACCGAAGAAATCCG...CGATCGCTTCATTA 964

     101 uSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaA 118
          |||         |||:::::::|||   ::::::|||
     965 ATCT.....GGCAATCAGCTACAACATCAATAT..... 992

     118 rgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGln 134
                                          |||:::|||
     993 ................................CATCACAGGAAGTAT 1007

     135 AspTrpGluGlnThrGlnLysIle...TyrAsnCysHisValLeuLeuAs 150
              :::|||   :::::::|||      :::|||::::::   |||
    1008 GCCGTTGATCAAAGAAGACGGATTGCTGTACAATGCCGGATTTCTTTGCC 1057

     150 nSerLys.GlyAlaValValAlaThrTyrArgLysThrHisLeuCysAsp 166
          :::::: |||   :::         |||:::|||   |||:::
    1058 GACGCGACGGA...ACTTACGAAATGTACGAAAAGCTCCATGTCACCCCG 1104

     167 ValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183
          ||||||   :::   |||                          :::
    1105 GACGAGATAAAGAGTTGGGGA.....................CTGAG 1130

     183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
          |||   :::      :::      :::|||   ::::::|||||||||:
    1131 CGGCGGCAAACAGCTTAAAACATTCGATACGGACTGTGCAAAGATAGGCA 1180

     200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
         ::   :::||||||||:::::::|||||||||||||||      :::|||
    1181 TACTGATCTGTTATGATGTGGAATTTCCGGAACTCTCCCGTCTGATGGCC 1230

     217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleTh 233
         :::   |||   :::||||||      |||              :::   ::
    1231 GACCAAGGAATGCAGATTCTGTTTGTACCGTTTCTCACCGATACACAAAA 1280

     233 rGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrG 250
         ::::   ::::::   :::|||      :::|||||||||||||||:::::
    1281 TGCTTATTCGCGTGTTCGGGTCTGCGCACAGGCACGTGCCATTGAGAACG 1330

     250 lnCysTyrValValAlaAlaAlaGlnCysGly.........ArgHisHis 263
         ::|||:::|||||| ~ |||::::::   |||         |||   |||
    1331 AATGCTTTGTGGTAATAGCCGGCAGTGTAGGCAATCTTCCCCGTGTGCAC 1380

     264 GluLysArgAlaSerTyrGlyHisSerMetValValAspPro........ 277
         :::      :::|||::::::|||   |||      |||
    1381 AATATGGATATTCAATATGCTCAGTCGGGAGTATTCACACCTTGCGATTT 1430

     278 ....TrpGlyThrValValAlaArgCysSerGluGlyProGly....... 290
             :::   |||         |||::::::      |||:::
    1431 CGCTTTTCCGACAGACGGAAAGCGTGCCGAAGCAACTCCGAATACAGAAA 1480

     291 ..LeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArg 306
           :::   ::::::   :::||||||:::   |||::::::|||:::
    1481 TGATCCTGGTTTCGGATGTAGATCTCGACTTATTGAACGAACTACACACT 1530

     307 HisLeuProVal......PheGlnHisArgArgProAspLeuTyr 319
         :::      |||      ::::::   ||||||   |||:::|||
    1531 TACGGCAGCGTTCGCAACCTGAAGGACAGGCGAAATGATGTATAT 1575

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-102-524-748

seq_documentation_block:
; Sequence 748, Application US/10102524
; GENERAL INFORMATION:
;  APPLICANT: Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Gaiger, Alexander
;  APPLICANT:  Gordon, Brian
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF KIDNEY CANCER
;  FILE REFERENCE: 210121.572
;  CURRENT APPLICATION NUMBER: US/10/102,524
;  CURRENT FILING DATE:  2002-03-19
;  NUMBER OF SEQ ID NOS: 1863
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 748
;   LENGTH: 270
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-102-524-748


alignment_scores:
           Quality:  129.50            Length:     59
             Ratio:  2.943               Gaps:      1
 Percent Similarity:  74.576   Percent Identity: 42.373

alignment_block:
 US-09-357-675C-21 x US-10-102-524-748   ..

 Align seg 1/1  to: US-10-102-524-748  from: 1  to: 270

     264 GluLysArgAlaSerTyr.........GlyHisSerMetValValAspPr 277
         :::   :::|||||||||         ||||||||   ||||||:::||
      21 GATGACAAAGCCTCCTATGTTGCCTGGGGACACAGCACCGTGGTGAACCC 70

     277 oTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuA 294
         |||||||   |||:::|||:::::::::         ::::::      :
```

```
     71 TTGGGGGGAGGTTCTAGCCAAAGCTGGCACAGAAGAAGCAATCGTGTATT 120

    294 laArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProVal 310
        ::   |||||||||:::   |||   ::::::|||::::::::::||||||
    121 CAGACATAGACCTGAAGAAGCTGGCTGAAATACGCCAGCAAATCCCCGTT 170

    311 PheGlnHisArgArgProAspLeuTyr 319
        |||::::::::::|||   |||||||||
    171 TTTAGACAGAAGCGATCAGACCTCTAT 197

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-209B-1148

seq_documentation_block:
; Sequence 1148, Application US/09540209B
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1001-001
;  CURRENT APPLICATION NUMBER: US/09/540,209B
;  CURRENT FILING DATE:  2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1148
;   LENGTH: 777
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-09-540-209B-1148


alignment_scores:
            Quality:  126.50            Length:    266
              Ratio:  0.923              Gaps:     12
 Percent Similarity:  51.504   Percent Identity: 27.444

alignment_block:
 US-09-357-675C-21 x US-09-540-209B-1148   ..

 Align seg 1/1  to: US-09-540-209B-1148  from: 1  to: 777

     59 AspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGlu..... 73
        :::||||||:::|||:::::            |||:::||||||
     46 AATAAACAGGAAAATCTCCGT............TTGCTCCGCGAAAAGCT 83

     74 ....AlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheA 89
                |||      :::   ::::::   |||||||||
     84 ATCACCTCTTCGCGGAACAACGGAGATTGTTGTTTTACCGGAGATG.... 129

     89 spPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeu 105
          |||   :::           ::::::     |||:::||||||:::
    130 ..TTTACAACAGGATTCAGCATGAACAGCCGGCTATTAGCCGAACCGGTT 177

    106 GlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLe 122
        :::|||      |||::::::      :::   |||   |||      ||
    178 TCCGGTACCACGCTCCGGAGTCTCAAAAATTATGCCATAGAATTTCATTT 227

    122 uTrpLeuSerLeuGlyGlyPhe......HisGluArgGlyGlnAspTrpG 137
        |   |||:::   |||:::|||      :::|||:::|||
    228 GTCATTGGCC...GGGAGTTTCATTTGTGAAGAACAAGGT.......... 264

    137 luGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGly 153
                  :::   ||||||         |||::::::     |||
    265 ..........TCTTATTATAACCGGGCTTTCCTGATCACTCCCGATGGA 303

    154 AlaValValAlaThrTyrArgLysThrHisLeuCysAspValGluIlePr 170
                       |||   |||   ||||||      :::
    304 CAGGAATTT...TACTATGACAAACGCCACCTCTTCCGCATG........ 342

    170 oGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerL 187
              |||      :::|||      |||         :::   |||
    343 .......GGACACGAAGCGGAACATTTTTCG.........GCAGGCAGCC 376

    187 euGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCys 203
          :::      :::         :::   :::|||   |||   ||||||

    377 GGAAAGTGATCATTCCCTACAATGGTTGGAACATCTGCCTGCAGGTATGT 426

    204 TyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAl 220
        ||||||:::|||||||||      |||      :::   :::
    427 TACGACCTCCGCTTTCCTGTCTGGAGCAGGAATGTGAACAATGAA...TA 473

    220 aGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaH 237
         ::::::|||   |||   :::::::::   :::            |||
    474 TGACCTCCTTATATATGTAGCCAGTTGGCCGACTCCACGTATTCAGGCA. 522

    237 isTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 253
          |||:::::::||||||   |||||||||||||||:::||||||||||||
    523 ..TGGAATACATTATTATGCGCACGTGCCATTGAAAATCAATGTTATGTA 570

    254 ValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGl 270
          ::::::   :::   |||:::     :::      :::       ||
    571 TGCGGTGTGAACCGTATAGGACAGGACGGCAACGGGCTCTGTTATCCGGG 620

    270 yHisSerMetValValAspProTrpGlyThrValValAlaArgCysSerG 287
        |:::|||   :::      |||   |||      :::|||         :
    621 GTATTCCGCTTTATATGGACCTAAAGGAGAAAACCTGGCAGGAACTCCCG 670

    287 luGly...ProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArg 302
        :::::        :::        :::::::::|||:::   |||
    671 ATTCGGAAGAAAAAATACAAACCATTGAACTTAGCCTGGAAGCCCTCACT 720

    303 GlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu 318
          :::|||:::   :::|||   ::::::         |||   |||
    721 ACTTTTCGTCATAAATTCCCTTGCTGGAAAGATGCAGACCCCTTTCTC 768

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-5721

seq_documentation_block:
; Sequence 5721, Application US/09540210B
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
;  FILE REFERENCE: PD-1037 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,210B
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/972,899
;  PRIOR FILING DATE: November 18, 1997
;  PRIOR APPLICATION NUMBER: 08/395,244
;  PRIOR FILING DATE: February 27, 1995
;  PRIOR APPLICATION NUMBER: 08/722,922
;  PRIOR FILING DATE: September 27, 1996
;  PRIOR APPLICATION NUMBER: 60/005,526
;  PRIOR FILING DATE: September 29, 1995
;  PRIOR APPLICATION NUMBER: 08/824,029
;  PRIOR FILING DATE: March 25, 1997
;  PRIOR APPLICATION NUMBER: 60/014,010
;  PRIOR FILING DATE: March 25, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/903,555
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/862,178
;  PRIOR FILING DATE: May 22, 1997
;  PRIOR APPLICATION NUMBER: 60/018,217
;  PRIOR FILING DATE: May 23, 1996
;  PRIOR APPLICATION NUMBER: 08/881,589
;  PRIOR FILING DATE: June 24, 1997
;  PRIOR APPLICATION NUMBER: 60/021,275
;  PRIOR FILING DATE: June 25, 1996
```

```
;  PRIOR APPLICATION NUMBER: 08/903,802
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/905,881
;  PRIOR FILING DATE: August 1, 1997
;  PRIOR APPLICATION NUMBER: 60/025,204
;  PRIOR FILING DATE: August 1, 1996
;  PRIOR APPLICATION NUMBER: 08/903,471
;  PRIOR FILING DATE: July 30, 1997
;  PRIOR APPLICATION NUMBER: 60/025,478
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/903,556
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/025,217
;  PRIOR FILING DATE: August 22, 1996
;  PRIOR APPLICATION NUMBER: 08/937,142
;  PRIOR FILING DATE: September 23, 1997
;  PRIOR APPLICATION NUMBER: 60/026,598
;  PRIOR FILING DATE: September 24, 1996
;  PRIOR APPLICATION NUMBER: 08/960,746
;  PRIOR FILING DATE: October 29, 1997
;  PRIOR APPLICATION NUMBER: 60/030,144
;  PRIOR FILING DATE: October 30, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/755,524
;  PRIOR FILING DATE: November 22, 1996
;  PRIOR APPLICATION NUMBER: 60/007,495
;  PRIOR FILING DATE: November 22, 1995
;  PRIOR APPLICATION NUMBER: 09/021,031
;  PRIOR FILING DATE: February 10, 1998
;  PRIOR APPLICATION NUMBER: 60/039,325
;  PRIOR FILING DATE: February 13, 1997
;  PRIOR APPLICATION NUMBER: 09/035,172
;  PRIOR FILING DATE: March 4, 1998
;  PRIOR APPLICATION NUMBER: 60/040,431
;  PRIOR FILING DATE: March 5, 1997
;  PRIOR APPLICATION NUMBER: 09/041,894
;  PRIOR FILING DATE: March 12, 1998
;  PRIOR APPLICATION NUMBER: 60/040,199
;  PRIOR FILING DATE: March 14, 1997
;  PRIOR APPLICATION NUMBER: 09/050,817
;  PRIOR FILING DATE: March 30, 1998
;  PRIOR APPLICATION NUMBER: 60/043,792
;  PRIOR FILING DATE: April 11, 1997
;  PRIOR APPLICATION NUMBER: 09/074,999
;  PRIOR FILING DATE: May 8, 1998
;  PRIOR APPLICATION NUMBER: 60/048,431
;  PRIOR FILING DATE: May 29, 1997
;  PRIOR APPLICATION NUMBER: 09/107,592
;  PRIOR FILING DATE: June 30, 1998
;  PRIOR APPLICATION NUMBER: 60/052,751
;  PRIOR FILING DATE: July 1, 1997
;  PRIOR APPLICATION NUMBER: 09/094,079
;  PRIOR FILING DATE: June 9, 1998
;  PRIOR APPLICATION NUMBER: 60/049,975
;  PRIOR FILING DATE: June 13, 1997
;  NUMBER OF SEQ ID NOS: 35654
;  SOFTWARE: PERL Program
; SEQ ID NO 5721
;   LENGTH: 252
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00212121
;   NAME/KEY: unsure
;   LOCATION: 65, 167
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-5721


alignment_scores:
            Quality:  124.00              Length:     71
              Ratio:  2.431                 Gaps:      2
 Percent Similarity:  71.831   Percent Identity: 38.028

alignment_block:
 US-09-357-675C-21 x US-09-540-210B-5721   ..

 Align seg 1/1  to: US-09-540-210B-5721  from: 1  to: 252

     252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
         ||||||::::::|||:::   :::         ::::::   :::|||||
       6 TATGTGGCCACAGCCTCTCCTGCC.........CGGGATGACAAAGCCTC 46

     268 rTyr.........GlyHisSerMetValValAspProTrpGlyThrValV 282
         ||||         |||   |||   ||||||:::|||||||||   |||:
      47 CTATGTTGCCTGGGGACANAGCACCGTGGTGAACCCTTGGGGGGAGGTTC 96

     282 alAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
         ::|||::::::::::       ::::::      :::   |||||||||
      97 TAGCCAAAGCTGGCACAGAAGAAGCAATCGTGTATTCAGACATAGACCTG 146

     299 AsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgAr 315
         :::   |||   ::::::|||::::::::::|||||||||::::::::||
     147 AAGAAGCTGGCTGAAATACGNCAGCAAATCCCCGTTTTTAGACAGAAGCG 196

     315 gProAspLeuTyr 319
         |   |||||||||
     197 ATCAGACCTCTAT 209

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-540-210B-5426

seq_documentation_block:
; Sequence 5426, Application US/09540210B
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
;  FILE REFERENCE: PD-1037 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,210B
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/972,899
;  PRIOR FILING DATE: November 18, 1997
;  PRIOR APPLICATION NUMBER: 08/395,244
;  PRIOR FILING DATE: February 27, 1995
;  PRIOR APPLICATION NUMBER: 08/722,922
;  PRIOR FILING DATE: September 27, 1996
;  PRIOR APPLICATION NUMBER: 60/005,526
;  PRIOR FILING DATE: September 29, 1995
;  PRIOR APPLICATION NUMBER: 08/824,029
;  PRIOR FILING DATE: March 25, 1997
;  PRIOR APPLICATION NUMBER: 60/014,010
;  PRIOR FILING DATE: March 25, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/903,555
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/862,178
;  PRIOR FILING DATE: May 22, 1997
;  PRIOR APPLICATION NUMBER: 60/018,217
;  PRIOR FILING DATE: May 23, 1996
;  PRIOR APPLICATION NUMBER: 08/881,589
;  PRIOR FILING DATE: June 24, 1997
```

```
;  PRIOR APPLICATION NUMBER: 60/021,275
;  PRIOR FILING DATE: June 25, 1996
;  PRIOR APPLICATION NUMBER: 08/903,802
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/023,308
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/905,881
;  PRIOR FILING DATE: August 1, 1997
;  PRIOR APPLICATION NUMBER: 60/025,204
;  PRIOR FILING DATE: August 1, 1996
;  PRIOR APPLICATION NUMBER: 08/903,471
;  PRIOR FILING DATE: July 30, 1997
;  PRIOR APPLICATION NUMBER: 60/025,478
;  PRIOR FILING DATE: July 31, 1996
;  PRIOR APPLICATION NUMBER: 08/903,556
;  PRIOR FILING DATE: July 31, 1997
;  PRIOR APPLICATION NUMBER: 60/025,217
;  PRIOR FILING DATE: August 22, 1996
;  PRIOR APPLICATION NUMBER: 08/937,142
;  PRIOR FILING DATE: September 23, 1997
;  PRIOR APPLICATION NUMBER: 60/026,598
;  PRIOR FILING DATE: September 24, 1996
;  PRIOR APPLICATION NUMBER: 08/960,746
;  PRIOR FILING DATE: October 29, 1997
;  PRIOR APPLICATION NUMBER: 60/030,144
;  PRIOR FILING DATE: October 30, 1996
;  PRIOR APPLICATION NUMBER: 08/826,847
;  PRIOR FILING DATE: April 10, 1997
;  PRIOR APPLICATION NUMBER: 60/015,533
;  PRIOR FILING DATE: April 10, 1996
;  PRIOR APPLICATION NUMBER: 08/755,524
;  PRIOR FILING DATE: November 22, 1996
;  PRIOR APPLICATION NUMBER: 60/007,495
;  PRIOR FILING DATE: November 22, 1995
;  PRIOR APPLICATION NUMBER: 09/021,031
;  PRIOR FILING DATE: February 10, 1998
;  PRIOR APPLICATION NUMBER: 60/039,325
;  PRIOR FILING DATE: February 13, 1997
;  PRIOR APPLICATION NUMBER: 09/035,172
;  PRIOR FILING DATE: March 4, 1998
;  PRIOR APPLICATION NUMBER: 60/040,431
;  PRIOR FILING DATE: March 5, 1997
;  PRIOR APPLICATION NUMBER: 09/041,894
;  PRIOR FILING DATE: March 12, 1998
;  PRIOR APPLICATION NUMBER: 60/040,199
;  PRIOR FILING DATE: March 14, 1997
;  PRIOR APPLICATION NUMBER: 09/050,817
;  PRIOR FILING DATE: March 30, 1998
;  PRIOR APPLICATION NUMBER: 60/043,792
;  PRIOR FILING DATE: April 11, 1997
;  PRIOR APPLICATION NUMBER: 09/074,999
;  PRIOR FILING DATE: May 8, 1998
;  PRIOR APPLICATION NUMBER: 60/048,431
;  PRIOR FILING DATE: May 29, 1997
;  PRIOR APPLICATION NUMBER: 09/107,592
;  PRIOR FILING DATE: June 30, 1998
;  PRIOR APPLICATION NUMBER: 60/052,751
;  PRIOR FILING DATE: July 1, 1997
;  PRIOR APPLICATION NUMBER: 09/094,079
;  PRIOR FILING DATE: June 9, 1998
;  PRIOR APPLICATION NUMBER: 60/049,975
;  PRIOR FILING DATE: June 13, 1997
;  NUMBER OF SEQ ID NOS: 35654
;  SOFTWARE: PERL Program
; SEQ ID NO 5426
;   LENGTH: 236
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00729617
US-09-540-210B-5426


alignment_scores:
           Quality:  118.50            Length:     79
             Ratio:  2.155               Gaps:      1
 Percent Similarity:  69.620   Percent Identity: 30.380

alignment_block:
 US-09-357-675C-21 x US-09-540-210B-5426   ..

 Align seg 1/1  to: US-09-540-210B-5426  from: 1  to: 236

     102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
         :::|||   :::   |||:::      ::::::   ::::::::::|||::
      10 GCAGAGAAAATTCCTGGTGAATCCACACAGAAGCTTTCTGAAGTAGCAAA 59

     118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
         :||||||::::::::::|||      |||::::::   |||
      60 GGAATGCAGCATATATCTCATTGGAGGCTCTATCCCTGAA.......... 99

     135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
           |||::::::   |||:::||||||      ::::::::::::
     100 .....GAGGATGCTGGGAAATTATATAACACCTGTGCTGTGTTTGGGCCT 144

     152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
         |||::::::::::|||   ||||||||   ||||||   |||:::::
     145 GATGGAACTTTACTAGCAAAGTATAGAAAGATCCATCTGTTTGACATTGA 194

     168 uIleProGlyGlnGlyProMetCysGluSerAsnSer 180
         ::::||||||:::      :::   ||||||:::::
     195 TGTTCCTGGAAAAATTACATTTCAAGAATCTAAAACA 231

seq_name: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:US-09-975-254-11297

seq_documentation_block:
; Sequence 11297, Application US/09975254
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15309)B
;  CURRENT APPLICATION NUMBER: US/09/975,254
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US/09/263,191
;  PRIOR FILING DATE: 1999-03-05
;  NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 11297
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 700904605H1
US-09-975-254-11297


alignment_scores:
           Quality:  107.00            Length:     79
             Ratio:  2.326               Gaps:      2
 Percent Similarity:  58.228   Percent Identity: 34.177

alignment_block:
 US-09-357-675C-21 x US-09-975-254-11297   ..

 Align seg 1/1  to: US-09-975-254-11297  from: 1  to: 264

     180 SerThrMetProGlyProSerLeuGluSerProValSerThrProAlaGl 196
         ||||||:::|||                        :::|||||:::||
      43 TCTACCATTCCA.................GTGTTTGAAACTCCCGTTGG 74

     196 yLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerL 213
         ||||||||||   |||:::|||::::::   |||:::|||   |||
      75 AAAAATAGGTGCTGCCATTTGTTGGGAGAACAGAATGCCACTGTTAAGGA 124
```

```
     213 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 229
     125 CAGCAATGTATGCCAAAGGTGTGGAGATATATTGTGCTCCAACAGCTGAT 174

     230 GlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAl 246
     175 TCCAGGGATGTT.....TGGCAAGCATCAATGACCCATATTGC 212

     246 aIleGluThrGlnCysTyrValValAlaAlaAlaGln 258
     213 ACTTGAAGGTGGATGTTTTGTTTTGTCAGCAAACCAG 249

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-115-123-90

seq_documentation_block:
; Sequence 90, Application US/10115123
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 94 Human Secreted Proteins
;  FILE REFERENCE: PZ029G30AP1D2
;  CURRENT APPLICATION NUMBER: US/10/115,123
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: PCT/US99/13418
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/089,507
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089,508
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089,509
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089,510
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/090,112
;  PRIOR FILING DATE: 1998-06-22
;  PRIOR APPLICATION NUMBER: 60/090,113
;  PRIOR FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 532
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
;   LENGTH: 1037
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-115-123-90


alignment_scores:
            Quality:   95.00            Length:    266
              Ratio:  0.856              Gaps:     16
 Percent Similarity:  41.729   Percent Identity: 24.060

alignment_block:
 US-09-357-675C-21 x US-10-115-123-90   ..

 Align seg 1/1  to: US-10-115-123-90  from: 1  to: 1037

      65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81
     182 AGGACATGTYGG.....GACAGCTCGATGCTCGGCCTGCTGCT 219

      81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluT 98
     220 GCTCTGCACCCCCAGGGCCTGG..... 241

      98 hrLeuHisLeuSerGluProLeu.....GlyGlyLysLeuLeuGluGlu 112
     242 ..CTCACCCTCTCTGGACCTGTCTGCTTCCAAGGAAGGGGACCCTCTGAG 289

     113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeu.....Se 125
     290 GTCCCACAG......AGGCCACCCCAGYTGTGGGTCGTGAGCATCTCTGT 333

     125 rLeuGlyGlyPheHis.....GluA 132
     334 CTTGCAGGGACAGCATCGTGGCCGAGCTGGACCGAGAGATGAGCAGGAGC 383

     132 rgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeu 148
     384 GTGGACGTGACCAA.....CACCAMCTT 406

     149 LeuAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCy 165
     407 .....CCTGCTCATGGCCGCCTCCATCTATC 432

     165 sAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrM 182
     433 TCCACGACCAGAACCCGGATGCCGCCCTGCGTGCGCTGCACCAGGGGGAC 482

     182 etProGly.....ProSerLeuGluSerProValSerThr 193
     483 AGCCTGGAGTGGTGAGTGGCCTCCCTGCTCTGGGCCAGCCC..... 523

     194 ProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGl 210
     524 .....AGGGAGGCAAGTGCCCCCTGCCACATCTCCAGGCTGCGCAC 564

     210 uLeuSerLeuAlaLeuAla.....GlnAlaGlyAlaGluIleL 223
     565 GGCCTCGCTGGCTGTCGTCATGGGAGCAGAGAAAGGTGGTGCTGAAATGA 614

     223 euThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGlu 239
     615 GGCCCTGGCCT.....GCTGTCCAGGCTCCAGCTCC..... 645

     240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAl 256
     646 .....CC 647

     256 aAlaGlnCysGlyArgHisHisGluLysArgAlaSer.....TyrG 270
     648 TGCCCAGTGTGGGAGGCACTCCCATCTGCGCACCAGGCTGCGGATCCAAG 697

     270 lyHisSerMetValValAspProTrpGlyThrValValAlaArgCysSer 286
     698 GACACGGT.....GCCCARGCTGCAACCCTCTGTTCC 729

     287 GluGlyProGly.....LeuCysLeuAlaArgIleAspLeuAsn 299
     730 CAAGGGCAGAGCAGAAAGCGGCTTTGTCTCTGCTCGGTTTCTGTGTCC 777

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-14411

seq_documentation_block:
; Sequence 14411, Application US/10105299
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et. al
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS950
;  CURRENT APPLICATION NUMBER: US/10/105,299
;  CURRENT FILING DATE:  2002-03-26
;  NUMBER OF SEQ ID NOS: 15197
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14411
;   LENGTH: 12280
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-105-299-14411


alignment_scores:
            Quality:   93.00            Length:    262
              Ratio:  0.838              Gaps:     15
 Percent Similarity:  42.366   Percent Identity: 24.427

alignment_block:
 US-09-357-675C-21 x US-10-105-299-14411   ..
```

```
Align seq 1/1  to: US-10-105-299-14411  from: 1  to: 12280

      65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81
         :::||||||           |||:::::::::   ||||||      ||
    4523 AGGACATGT...........CGGGACAGCTCGATGCTCGGCCTGCTGCT 4560

      81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluT 98
         |         |||:::|||:::
    4561 GCTCTGCACCCCCAGGGCCTGG.......................... 4582

      98 hrLeuHisLeuSerGluProLeu......GlyGlyLysLeuLeuGluGlu 112
           |||   ||||||   |||:::        |||:::      :::|||
    4583 ..CTCACCCTCTCTGGACCTGTCTGCTTCCAAGGAAGGGGACCCTCTGAG 4630

     113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeu...........Se 125
              |||      |||         ||||||:::
    4631 GTCCCACAG......AGGCCACCCCAGCTGTGGGTCGTGAGCATCTCTGT 4674

     125 rLeuGlyGlyPheHis.............................GluA 132
         |||   |||   |||                               ||||
    4675 CTTGCAGGGACAGCATCGTGGCCGAGCTGGACCGAGAGATGAGCAGGAGC 4724

     132 rgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeu 148
         |||||:::|||   ::::::
    4725 GTGGACGTGACCAACACCAC.......................... 4744

     149 LeuAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCy 165
                        :::   ::::::   :::   ||||||
    4745 ..................CTTCCTGCTCATGGCCGCCTCCATCTATC 4773

     165 sAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrM 182
                   ||||||      |||   |||   :::   :::   :
    4774 TCCACGACCAGAACCCGGATGCCGCCCTGCGTGCGCTGCACCAGGGGGAC 4823

     182 etProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIle 198
         ::||||||              |||:::.     ||||||   :::
    4824 AGCCTGGAGTGGTGAGTGGCCTCCCTGCTCTGGGGCCAGCC...CAGGGA 4870

     199 Gly.LeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaL 215
         |||   |||   |||:::     |||:::        |||||||||:
    4871 GGCAAGTGCCCCCTGCCACATCTCCAGGCTGCGCACGGCCTCGCTGGCTG 4920

     215 euAla...........GlnAlaGlyAlaGluIleLeuThrTyrProSer 227
         :::::           ::::::||||||||||:::      :::|||
    4921 TCGTCATGGGAGCAGAGAAAGGTGGTGCTGAAATGAGGCCCTGGCCT... 4967

     228 AlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAl 244
                :::::   :::||||||
    4968 .........GCTGTCCAGGCTCCAGCTCC..................... 4987

     244 aArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyA 261
                                      ||||||||||||
    4988 ................................CCTGCCCAGTGTGGGA 5003

     261 rgHisHisGluLysArgAlaSer.........TyrGlyHisSerMetVal 274
         |||||   :::   |||:::              ||||||:::
    5004 GGCACTCCCATCTGCGCACCAGGCTGCAGATCCAAGGACACGGT...... 5047

     275 ValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGly.. 290
                 :::   :::::::::   ||||||:::|||   :::
    5048 ............GCCCAGGCTGCAACCCTCTGTTCCCAAGGGCAGAGCAG 5085

     291 .......LeuCysLeuAlaArgIleAspLeuAsn 299
                |||||||||:::   ::::::::::::
    5086 AAAGCGGCTTTGTCTCTGCTCGGTTTCTGTGTCC 5119

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-14786

seq_documentation_block:
; Sequence 14786, Application US/10105299
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et. al
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS950
;  CURRENT APPLICATION NUMBER: US/10/105,299
;  CURRENT FILING DATE:  2002-03-26
;  NUMBER OF SEQ ID NOS: 15197
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14786
;   LENGTH: 12280
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-105-299-14786


alignment_scores:
            Quality:   93.00            Length:    262
              Ratio:  0.838              Gaps:     15
 Percent Similarity:  42.366   Percent Identity: 24.427

alignment_block:
 US-09-357-675C-21 x US-10-105-299-14786   ..

 Align seq 1/1  to: US-10-105-299-14786  from: 1  to: 12280

      65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81
         :::||||||           |||:::::::::   ||||||      ||
    4523 AGGACATGT...........CGGGACAGCTCGATGCTCGGCCTGCTGCT 4560

      81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluT 98
         |         |||:::|||:::
    4561 GCTCTGCACCCCCAGGGCCTGG.......................... 4582

      98 hrLeuHisLeuSerGluProLeu......GlyGlyLysLeuLeuGluGlu 112
           |||   ||||||   |||:::        |||:::      :::|||
    4583 ..CTCACCCTCTCTGGACCTGTCTGCTTCCAAGGAAGGGGACCCTCTGAG 4630

     113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeu...........Se 125
              |||      |||         ||||||:::
    4631 GTCCCACAG......AGGCCACCCCAGCTGTGGGTCGTGAGCATCTCTGT 4674

     125 rLeuGlyGlyPheHis.............................GluA 132
         |||   |||   |||                               ||||
    4675 CTTGCAGGGACAGCATCGTGGCCGAGCTGGACCGAGAGATGAGCAGGAGC 4724

     132 rgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeu 148
         |||||:::|||   ::::::
    4725 GTGGACGTGACCAACACCAC.......................... 4744

     149 LeuAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCy 165
                        :::   ::::::   :::   ||||||
    4745 ..................CTTCCTGCTCATGGCCGCCTCCATCTATC 4773

     165 sAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrM 182
                   ||||||      |||   |||   :::   :::   :
    4774 TCCACGACCAGAACCCGGATGCCGCCCTGCGTGCGCTGCACCAGGGGGAC 4823

     182 etProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIle 198
         ::||||||              |||:::      ||||||   :::
    4824 AGCCTGGAGTGGTGAGTGGCCTCCCTGCTCTGGGGCCAGCC...CAGGGA 4870

     199 Gly.LeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaL 215
         |||   |||   |||:::     |||:::        |||||||||:
    4871 GGCAAGTGCCCCCTGCCACATCTCCAGGCTGCGCACGGCCTCGCTGGCTG 4920

     215 euAla...........GlnAlaGlyAlaGluIleLeuThrTyrProSer 227
         :::::           ::::::||||||||||:::      :::|||
    4921 TCGTCATGGGAGCAGAGAAAGGTGGTGCTGAAATGAGGCCCTGGCCT... 4967

     228 AlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAl 244
                :::::   :::||||||
    4968 .........GCTGTCCAGGCTCCAGCTCC..................... 4987

     244 aArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyA 261
```

```
                                                 ||||||||||||||
    4988 ..................................CCTGCCCAGTGTGGGA 5003

     261 rgHisHisGluLysArgAlaSer.........TyrGlyHisSerMetVal 274
         |||||   :::   |||:::              ||||||:::
    5004 GGCACTCCCATCTGCGCACCAGGCTGCAGATCCAAGGACACGGT...... 5047

     275 ValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGly.. 290
                  :::   :::::::::   ||||||:::|||   :::
    5048 ............GCCCAGGCTGCAACCCTCTGTTCCCAAGGGCAGAGCAG 5085

     291 .......LeuCysLeuAlaArgIleAspLeuAsn 299
              ||||||||||:::   ::::::::::::
    5086 AAAGCGGCTTTGTCTCTGCTCGGTTTCTGTGTCC 5119

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-108

seq_documentation_block:
; Sequence 108, Application US/10105299
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et. al
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS950
;  CURRENT APPLICATION NUMBER: US/10/105,299
;  CURRENT FILING DATE:  2002-03-26
;  NUMBER OF SEQ ID NOS: 15197
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
;   LENGTH: 1536
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-105-299-108


alignment_scores:
           Quality:   92.50            Length:    326
             Ratio:  0.638              Gaps:     16
 Percent Similarity:  44.479   Percent Identity: 24.233

alignment_block:
 US-09-357-675C-21 x US-10-105-299-108   ..

 Align seq 1/1  to: US-10-105-299-108  from: 1  to: 1536

       6 ThrArgProPro...HisArgPheLeuSerLeuLeuCysProGlyLeuAr 21
         |||   ||||||   |||                  |||:::|||
      13 ACATCTCCCCCATAGCAC.....................CCTGCCCTCAT 41

      21 gIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetA 38
             |||   |||                  |||:::|||   |||::::
      42 GGGACCTGCCCTC.................CCTCAGCCGTCAGCCATCA 73

      38 laIleSerSerSerSerCysGluLeuProLeuVal.............. 49
         ::   :::     |||::::::   |||||:|:::
      74 GCCATGGCCCTCCCAGTGCCTCCTAGCCCCTTCTTCCAAGGAGCAGAGAG 123

      50 ..............AlaValCysGlnValThrSerThrProAspLys.. 60
                       |||:::      :::   |||   |||   |||
     124 GTGGCCACCGGGGGTGCTCTGTCCTACCTCCACTCTCTGCCCCTAAAGAT 173

      61 .........GlnGlnAsnPheLysThrCysAlaGluLeuValArgGluA 74
                 |||::::::::::      :::::::|||            |
     174 GGGAGGAGACCAGCGGTCCATGGGTCTGGCCTGTGAGTCTCCCCTTGCAG 223

      74 laAlaArgLeuGlyAlaCysLeuAlaPhe...LeuProGluAlaPheAsp 89
         ||      ||||||         |||:::   |||          :::
     224 CCTGGTCACTAGGCATCACCCCCGCTTTGGTTCTTCAGATGCTCTTGGGG 273

      90 PheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGl 106
         ||||||:::      |||::::::::      |||:::   |||
     274 TTCATAGGGGCAGGTCCTAGTCGGGCAGGGCCCCTGACCCTCCCG..... 318

     106 yGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuT 123
                                                          |
     319 ..................................................GCCT 322

     123 rpLeuSer.....................LeuGlyGlyPheHisGluArg 132
         |||||                        :::      |||   ||||||
     323 GGCTTCACTCTCCCTGACGGCTGCCATTGGTCCACCCTTTCATAGAGAGG 372

     133 GlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLe 149
                                                 :::|||||
     373 ....................................CCTGCTTTGTT 383

     149 uAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysA 166
         |:::|||:::|||         |||:::      :::|||:::      :
     384 ACAAAGCTCGGGTCTCCCTCCTGCAGCTCGGTTAAGTACCCGAGGCCTCT 433

     166 spValGluIleProGly.GlnGlyProMetCysGluSerAsnSerThrMe 182
         ::   ::::::   |||    ||||||   ::::::      :::|||::
     434 CTTAAGATGTCCAGGGCCCCAGGCCCGCGGGCACAGCCAGCCCAAACCTT 483

     182 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199
         :   ||||||:::   :::|||      |||      ::::::      |
     484 G...GGCCCTGGAAGATCCTCCACCCCATCACTAGAGTGCTCTGACCCTG 530

     199 lyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeu.AlaLe 215
         ||||||:::   :::          |||:::|||:::|||||||||    ||
     531 GGCTTTCACGGGCCCCATTCCACCGCCTCCCCAACTTGAGCCTGTGACCT 580

     215 uAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAla..... 228
         |:::      |||:::   :::::::::      ||||||
     581 TGGGACCAAAGGGGGAGTCCTCGTCTCTTGTGACTCAGCAGAGGCAGTGG 630

     229 .........................PheGlySerIleThrGlyProAla 236
                                  ||||||         ||||||
     631 CCACGTTCAGGAGGGGCCGGTGGCCTGGAGGCTCAGCCCACCCTCCAGCT 680

     237 HisTrpGluValLeuLeuArgAlaArgAlaIleGlu.....Th 249
             :::   :::|||||||::::::   :::|||      ::
     681 TTTCCTCAGGGTGTCCTGAGGTCCAAGATTCTGGAGCAATCTGACCCTTC 730

     249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
         :         :::::::::|||:::||||||         |||:::
     731 TCCAAAGGCTCTGTTATCAGCTGGGCAGTGC.........CACCAA.... 767

     266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
                       |||::::::::::   |||   ||||||   :::
     768 .................TCCCTGGCCATTTGGCCCCAGGGGACGTGGGCC 800

     283 AlaArgCysSerGluGlyProGly 290
         :::|||   :::||||||   |||
     801 TGCAGGCTGCAGGAGGGCACTGGA 824

seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-1860

seq_documentation_block:
; Sequence 1860, Application US/10105299
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et. al
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS950
;  CURRENT APPLICATION NUMBER: US/10/105,299
;  CURRENT FILING DATE:  2002-03-26
;  NUMBER OF SEQ ID NOS: 15197
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1860
;   LENGTH: 2102
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-105-299-1860


alignment_scores:
```

```
          Quality:   92.50            Length:    232
            Ratio:  0.881              Gaps:     12
Percent Similarity:  45.259   Percent Identity: 19.828

alignment_block:
US-09-357-675C-21 x US-10-105-299-1860   ..

Align seg 1/1  to: US-10-105-299-1860  from: 1  to: 2102

     55 ThrSerThrProAspLysGln.....................GlnAs 63
        |||:::||||||   :::|||                       :::||
    214 ACAGAAACACCAGTTTCTCAGGAGGATGCCTTGAATCTCATGAACGAGAA 263

     63 nPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaC 80
        |:::                 :::::::::||||||:::   ||||||
    264 TATAGACATTCTGGAGACAGCGATCAAGCAGGCAGCTGAGCAGGGTGCTC 313

     80 ysLeuAlaPheLeuProGlu.............................. 86
          :::       ||||||
    314 GAATCATTGTGACTCCAGAAGATGCACTTTATGGATGGAAATTTACCAGG 363

     87 ......AlaPheAspPheIle............................. 91
             :::|||   ::::::
    364 GAAACTGTTTTCCCTTATCTGGAGGATATCCCAGACCCTCAGGTGAACTG 413

     92 .......AlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 106
             ::::::||||||   :::      |||         |||:::
    414 GATTCCGTGTCAAGACCCCCACAGATTTGGTCAC......ACACCAGTAC 457

    106 lyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeu 122
          :::::::|||            :::   ||||||::::::   ::::::
    458 AAGCAAGACTC............AGCTGCCTGGCCAAGGACAACTCTATC 495

    123 TrpLeu......SerLeuGly............................. 127
        ::::::      :::||||||
    496 TATGTCTTGGCAAATTTGGGGGACAAAAAGCCATGTAATTCCCGTGACTC 545

    128 ................GlyPheHisGluArgGlyGlnAspTrpGluGlnT 139
                       |||:::   :::
    546 CACGTGTCCTCCTAATGGCTACTTTCAA..................... 573

    139 hrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaVal 155
                |||||   :::|||:::   |||::::::|||   :::
    574 ...........TACAATACCAATGTGGTGTATAATACAGAAGGAAAACTC 612

    156 ValAlaThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGl 172
        ||||||   |||:::|||   ||||||
    613 GTGGCACGTTACCATAAGTACCACCTG....................... 639

    172 nGlyProMetCysGluSerAsnSerThrMetProGlyProSerLeuGluS 189
                  |||   :::   ::::::|||            |||:
    640 .......TACTCTGAGCCTCAGTTTAATGTCCCT...........GAAA 670

    189 erPro...........ValSerThrProAlaGlyLysIleGlyLeuAla 201
        ::|||           :::|||      |||::::::|||:::
    671 AGCCGGAGTTGGTGACTTTCAACACCGCATTTGGAAGGTTTGGCATTTTC 720

    202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGln.. 217
        :::|||:::|||:::   |||   :::   ::::::::::|||::::::
    721 ACGTGCTTTGATATATTCTTCTATGATCCTGGTGTTACCCTGGTGAAAGA 770

    218 .AlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle 232
           ::::::   :::   :::|||:::|||:::   ::::::
    771 TTTCCATGTGGACACCATACTGTTTCCCACAGCTTGGATGAACGTT 816
```

```
OM of: US-09-357-675C-21 to: EST:*   out_format : pfs

Date: Apr 29, 2002 10:01 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1h
 -Q=/cgn2_1/USPTO_spool/US09357675/runat_29042002_091756_3146/app_query.fasta_1.1208
 -DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09357675_@CGN1_1_6497
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 2545.690000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_htc:AK004988         + 1479.00 2490.04  1.9e-129  2993  ! AK004988 Mus musculus adult ma
gb_est1:AL520768        + 1458.00 2468.05  3.2e-128   890  ! AL520768 AL520768 LTI_NFL004_NE
gb_htc:AK002269         + 1458.00 2463.11  6.0e-128  1374  ! AK002269 Mus musculus adult ma
gb_est1:AL529152        + 1413.00 2391.96  5.5e-124   844  ! AL529152 AL529152 LTI_NFL001_NE
gb_est2:BI822844        + 1357.50 2296.29  1.2e-118   928  ! BI822844 603040166F1 NIH_MGC_11
gb_est2:BI769604        + 1308.00 2212.98  5.1e-114   846  ! BI769604 603054985F1 NIH_MGC_12
gb_est2:BG436916        + 1298.00 2196.67  4.1e-113   793  ! BG436916 602488424F1 NIH_MGC_18
gb_est1:AL559163        + 1253.00 2119.04  8.7e-109   861  ! AL559163 AL559163 LTI_NFL008_TC
gb_est2:BG762506        + 1218.00 2060.75  1.5e-105   764  ! BG762506 602733954F1 NIH_MGC_49
gb_est2:BG703176        + 1206.00 2039.92  2.2e-104   790  ! BG703176 602686082F1 NIH_MGC_95
gb_est1:AL520767        - 1201.50 2029.66  8.3e-104   992  ! AL520767 AL520767 LTI_NFL004_NE
gb_est1:AL522373        - 1179.50 1992.54  9.7e-102   960  ! AL522373 AL522373 LTI_NFL004_NE
gb_est2:BI757823        + 1169.00 1974.81  9.5e-101   946  ! BI757823 603030404F1 NIH_MGC_11
gb_est2:BI755493        + 1135.50 1919.44  1.1e-97    813  ! BI755493 603027349F1 NIH_MGC_11
gb_est2:BI711300        - 1115.00 1887.49  6.9e-96    625  ! BI711300 id96f12.x1 Human insul
gb_est2:BI752623        + 1109.00 1875.25  3.3e-95    746  ! BI752623 603028471F1 NIH_MGC_11
gb_est2:BE895409        + 1097.50 1854.47  4.8e-94    828  ! BE895409 601434778F1 NIH_MGC_72
gb_est2:BI647584        + 1089.00 1839.92  3.1e-93    832  ! BI647584 603278164F1 NIH_CGAP_M
gb_est2:BG776560        + 1087.00 1837.09  4.4e-93    791  ! BG776560 602663693F1 NIH_MGC_59
gb_est2:BG532265        + 1079.50 1824.45  2.2e-92    781  ! BG532265 602561355F1 NIH_MGC_61
gb_est2:BG963013        + 1077.00 1818.98  4.5e-92    869  ! BG963013 602828020F1 NCI_CGAP_C
gb_est1:AL562604        - 1067.50 1803.00  3.5e-91    853  ! AL562604 AL562604 LTI_NFL003_NE
gb_est2:BF968272        + 1045.50 1764.89  4.7e-89    900  ! BF968272 602269658F1 NIH_MGC_84
gb_est2:BI332807        + 1041.50 1758.55  1.0e-88    863  ! BI332807 602985469F1 NCI_CGAP_L
gb_est2:BE853807        + 1035.00 1750.91  2.8e-88    638  ! BE853807 ux23g06.y1 Soares_thym
gb_est2:BG667049        + 1025.00 1730.04  4.1e-87    893  ! BG667049 DRABTC11 Rat DRG Libra
gb_est2:BF696074        + 1013.50 1711.32  4.5e-86    827  ! BF696074 601852067F1 NIH_MGC_56
gb_est2:BI327600        + 1006.50 1699.75  2.0e-85    801  ! BI327600 602979783F1 NCI_CGAP_L
gb_est2:BI711712        + 1006.00 1702.59  1.4e-85    579  ! BI711712 id96f12.y1 Human insul
gb_est1:AL580840        - 1003.00 1693.38  4.5e-85    830  ! AL580840 AL580840 LTI_NFL008_TC
gb_est2:BI916460        + 1000.50 1691.78  5.5e-85    657  ! BI916460 603182527F1 NIH_MGC_12
gb_est2:BG432389        +  999.50 1686.49  1.1e-84    900  ! BG432389 602495419F1 NIH_MGC_75
gb_est2:BI340308        +  952.00 1610.92  1.8e-80    561  ! BI340308 365533 MARC 2PIG Sus s
gb_est2:BF780562        +  951.50 1605.73  3.4e-80    821  ! BF780562 602103103F1 NCI_CGAP_K
gb_est2:BE392424        +  939.00 1586.71  3.9e-79    672  ! BE392424 601308982F1 NIH_MGC_44
gb_est2:BF322101        +  931.00 1574.50  1.9e-78    593  ! BF322101 maa34b02.y1 NCI_CGAP_L
gb_est2:BM142111        +  912.00 1542.93  1.1e-76    552  ! BM142111 if26h04.y1 Melton Norm
gb_est2:BF385518        +  903.00 1523.52  1.3e-75    789  ! BF385518 602047063F1 NCI_CGAP_L
gb_est2:BM142010        +  896.00 1515.97  3.4e-75    537  ! BM142010 if25d10.y1 Melton Norm
gb_est2:BF135684        +  896.00 1509.60  7.7e-75    940  ! BF135684 601780826F1 NCI_CGAP_L
gb_est1:AL547581        +  894.00 1507.01  1.1e-74    875  ! AL547581 AL547581 LTI_NFL006
gb_est2:BF243444        +  889.50 1499.48  2.8e-74    864  ! BF243444 601878153F1 NIH_MGC
gb_est2:BE287010        +  887.50 1499.24  2.9e-74    654  ! BE287010 601097244F1 NCI_CGA
gb_est1:AW956706        +  879.00 1486.20  1.6e-73    576  ! AW956706 EST368776 MAGE rese
gb_est2:BI341863        +  862.00 1458.24  5.6e-72    527  ! BI341863 369447 MARC 2PIG Su

seq_name: gb_htc:AK004988

seq_documentation_block:
LOCUS       AK004988               2993 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300013F05:nitrilase 1, full insert sequence.
ACCESSION   AK004988
VERSION     AK004988.1  GI:12836590
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1300013F05.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2993)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
```

```
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .2993
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="MGD:MGI:1895066"
                      /db_xref="taxon:10090"
                      /clone="1300013F05"
                      /sex="male"
                      /tissue_type="liver"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     gene             169. .1041
                      /gene="Nit1"
     CDS              169. .1041
                      /gene="Nit1"
                      /note="data source:MGD, source key:MGI:1350916,
                      evidence:ISS
                      nitrilase 1
                      putative"
                      /codon_start=1
                      /protein_id="BAB23723.1"
                      /db_xref="GI:12836591"
                      /translation="MSSSTSWELPLVAVCQVTSTPNKQENFKTCAELVQEAARLGACL
                      AFLPEAFDFIARNPAETLLLSEPLNGDLLGQYSQLARECGIWLSLGGFHERGQDWEQN
                      QKIYNCHVLLNSKGSVVASYRKTHLCDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGK
                      VGLAICYDMRFPELSLKLAQAGAEILTYPSAFGSVTGPAHWEVLLRARAIESQCYVIA
                      AAQCGRHHETRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLHFLQQMRQHLPVFQH
                      RRPDLYGSLGHPLS"
     polyA_signal     2973. .2978
                      /note="putative"
     polyA_site       2993
                      /note="putative"
BASE COUNT      785 a    739 c    747 g    722 t
ORIGIN


alignment_scores:
            Quality: 1479.00            Length:    326
              Ratio:  4.833               Gaps:      2
 Percent Similarity:  93.865   Percent Identity: 84.356

alignment_block:
 US-09-357-675C-21 x AK004988   ..

 Align seg 1/1  to: AK004988  from: 1  to: 2993

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         ||||||||||||||||||||||||||:::         |||||||||
      73 CTGGGCTTCATCACCAGGCCTCCTCACCAA.........CTCCTGTGTAC 113

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
          |||   |||:::   :::      |||||||||:::|||||||||||||
     114 CGGATACCGATTACTTCGAACCCCAGTACTTTGTACTCAGCCCAGGCCCA 163

      35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
         ||:::|||:::   ||||||:::|||   ||||||||||||||||||||

     164 GAACCATGTCC...TCATCAACTTCCTGGGAGCTGCCCCTGGTGGCTGTG 210

      52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
         ||||||||||||||||||||||:::||||||:::||||||||||||||||
     211 TGCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACATGTGC 260

      68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
         ||||||||||:::||||||||||||||||||||||||||||||||||||
     261 TGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCCTTTCTGC 310

      85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
         |||||||||||||||||||||||||:::|||||||||||||||   |||
     311 CTGAGGCATTTGACTTTATTGCACGAAACCCTGCCGAGACATTACTCCTG 360

     102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
         ||||||||||||:::|||   ||||||   :::|||:::||||||||||
     361 TCCGAACCACTGAATGGGGATCTTTTGGGCCAATATAGCCAGCTTGCCAG 410

     118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
         ||||||||||:::||||||||||||||||||||||||||||||||||||
     411 GGAATGTGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCCAAG 460

     135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
         ||||||||||:::||||||||||||||||||||||||||||||||||||
     461 ACTGGGAGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAACAGC 510

     152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGl 168
         ||||||:::|||||||||:::||||||||||||||||||||||||||||
     511 AAGGGATCAGTAGTGGCCAGTTACAGGAAGACACATCTGTGCGATGTAGA 560

     168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
         ||||||||||||||||||||   |||||||||   |||   ||||||
     561 GATCCCAGGTCAGGGGCCGATGAGAGAAAGCAACTATACCAAGCCTGGAG 610

     185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
           :::||||||   ||||||:::||||||||||||||:::|||||||||
     611 GCACTCTTGAGCCACCTGTCAAGACACCGGCTGGCAAGGTTGGTCTAGCA 660

     202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
         :::||||||||||||||||||||||||||||||||   ||||||||||
     661 ATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTCAAGC 710

     218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
         ||||||||||||||||||||||||||||||||||||||:::|||||||
     711 TGGGGCAGAAATACTTACTTATCCTTCAGCCTTTGGATCTGTTACAGGTC 760

     235 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
         ||||||||||||||||||||||||||||||||||||||||||:::||||||
     761 CAGCCCACTGGGAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAGTGC 810

     252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
         ||||||:::||||||||||||||||||||||||||||   ||||||||
     811 TATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGCAAG 860

     268 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     861 TTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCT 910

     285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
         ||||||||||||||||||||||||||||||||||||||||||:::::|||
     911 GCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTTCTA 960

     302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
         :::|||:::|||:::||||||||||||||||||||||||||||||||||
     961 CAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGACCT 1010

     318 uTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||:::||||||||||||||||||
    1011 CTATGGCAGTCTGGGTCATCCACTCTCT 1038

seq_name: gb_est1:AL520768

seq_documentation_block:
```

```
LOCUS       AL520768                890 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
            prime, mRNA sequence.
ACCESSION   AL520768
VERSION     AL520768.1  GI:12784261
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      191 a    251 c    241 g    206 t      1 others
ORIGIN


alignment_scores:
            Quality: 1458.00            Length:    277
              Ratio:  5.302               Gaps:      0
 Percent Similarity:  99.278   Percent Identity: 99.278

alignment_block:
 US-09-357-675C-21 x AL520768   ..

 Align seg 1/1  to: AL520768  from: 1  to: 890

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         |||||||||  ||||||||||||||||||||||||||||||||||||||||
      60 ATGCTGGGCGTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 109

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     110 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 159

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     160 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 209

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     210 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 259

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     260 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 309

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     310 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 359

     101 LeuSerGluProLeuGlyGlyLysLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     360 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 409

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     410 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 459

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     460 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 509

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     510 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 559

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     560 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 609

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     610 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 659

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     660 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 709

     217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     710 AGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAG 759

     234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     760 GCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCGAAACCCAG 809

     251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     810 TGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 859

     267 aSerTyrGlyHisSerMetValValAspPro 277
         ||||||||||||||||||||  ||||||
     860 AAGTTATGGCCACAGCATGGTGGKAGACCCC 890

seq_name: gb_htc:AK002269

seq_documentation_block:
LOCUS       AK002269                1374 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:0610007B07:nitrilase 1, full insert sequence.
ACCESSION   AK002269
VERSION     AK002269.1  GI:12832130
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:0610007B07.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
```

```
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1374)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0. Second strand cDNA
            was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1374
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1891931"
                     /db_xref="taxon:10090"
                     /clone="0610007B07"
                     /sex="male"
                     /tissue_type="kidney"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     gene            1. .1374
                     /gene="Nit1"
     misc_feature    1. .1374
                     /gene="Nit1"
                     /note="data source:MGD, source key:MGI:1350916,
                     evidence:ISS
                     nitrilase 1"
BASE COUNT      330 a    359 c    342 g    343 t
ORIGIN


alignment_scores:
           Quality: 1458.00            Length:    327
             Ratio:  4.780              Gaps:      3
 Percent Similarity:  93.272   Percent Identity: 83.792

alignment_block:
 US-09-357-675C-21 x AK002269   ..

 Align seg 1/1  to: AK002269  from: 1  to: 1374

       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
         |||||||||||||||||||||||||||:::         |||||||||
     119 CTGGGCTTCATCACCAGGCCTCCTCACCAA.........CTCCTGTGTAC 159

      18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
          |||   |||:::   :::      ||||||||||:::|||||||||||
     160 CGGATACCGATTACTTCGAACCCCAGTACTTTGTACTCAGCCCAGGCCCA 209

      35 rgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
         ||:::|||:::   ||||||:::|||   ||||||||||||||||||||
     210 GAACCATGTCC...TCATCAACTTCCTGGGAGCTGCCCCTGGTGGCTGTG 256

      52 CysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAl 68
         ||||||||||||||||||:::|||||||:::|||||||||||||||||||
     257 TGCCAGGTAACATCAACACCAAACAAGCAAGAGAACTTTAAAACATGTGC 306

      68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
     307 TGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCCTTTCTGC 356

      85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
         |||||||||||||||||||||||||:::|||||||||||||||   |||
     357 CTGAGGCATTTGACTTTATTGCACGAAACCCTGCCGAGACATTACTCCTG 406

     102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaAr 118
         ||||||||||||:::|||   ||||||   :::|||:::||||||||||
     407 TCCGAACCACTGAATGGGGATCTTTTGGGCCAATATAGCCAGCTTGCCAG 456

     118 gGluCys.GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGln 134
         |||||||  |||:::|||||||||||||||||||||||||||||||||||
     457 GGAATGTTGGAATCTGGCTGTCCTTGGGCGGTTTCCACGAGCGTGGCCAA 506

     135 AspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSe 151
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
     507 GACTGGGAGCAGAATCAGAAAATCTACAATTGTCATGTGCTTTTGAACAG 556

     151 rLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValG 168
         |||||||:::|||||||||:::|||||||||||||||||||||||||||||
     557 CAAGGGATCAGTAGTGGCCAGTTACAGGAAGACACATCTGTGCGATGTAG 606

     168 luIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGly 184
         ||||||||||||||||||||||   |||||||||   |||   ||||||
     607 AGATCCCAGGTCAGGGGCCGATGAGAGAAAGCAACTATACCAAGCCTGGA 656

     185 ProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
            :::||||||   ||||||:::|||||||||||||||:::||||||||
     657 GGCACTCTTGAGCCACCTGTCAAGACACCGGCTGGCAAGGTTGGTCTAGC 706

     201 aValCysTyrAspMetArgProProGluLeuSerLeuAlaLeuAlaGlnA 218
         |:::|||||||||||||||||||||||||||||||   ||||||||||
     707 AATCTGTTATGACATGCGGTTCCCTGAACTTTCTTTGAAATTGGCTCAAG 756

     218 laGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGly 234
         ||||||||||||||||||||||||||||||||||   |||:::||||||
     757 CTGGGGCAGAAATACTTACTTATCCTTCAGCCTTTAGATCTGTTACAGGT 806
```

```
     235 ProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCy 251
         ||||||||||||||||||||||||||||||||||||||||||||:::|||||
     807 CCAGCCCACTGGGAGGTGCTGCTGCGGGCCCGCGCCATTGAATCTCAGTG 856

     251 sTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaS 268
         |||||||:::|||||||||||||||||||||||||||||   |||||||
     857 CTATGTAATAGCAGCAGCGCAGTGTGGACGCCACCATGAAACAAGAGCAA 906

     268 erTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArg 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     907 GTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGC 956

     285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLe 301
         ||||||||||||||||||||||||||||||||||||||||||||:::::||
     957 TGCTCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTCCACTTTCT 1006

     301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318
         |:::|||:::|||:::|||||||||||||||||||||||||||||||||
    1007 ACAACAGATGCGCCAACACCTGCCTGTGTTCCAGCACCGCAGACCTGACC 1056

     318 euTyrGlyAsnLeuGlyHisProLeuSer 327
         ||||||||:::|||||||||||||||||||
    1057 TCTATGGCAGTCTGGGTCATCCACTCTCT 1085

seq_name: gb_est1:AL529152

seq_documentation_block:
LOCUS       AL529152                844 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5
            prime, mRNA sequence.
ACCESSION   AL529152
VERSION     AL529152.1  GI:12792645
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DD004YM06"
                     /clone_lib="LTI_NFL001_NBC4"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      182 a    238 c    235 g    189 t
ORIGIN


alignment_scores:
            Quality: 1413.00            Length:    263
              Ratio:  5.373              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-357-675C-21 x AL529152   ..

 Align seg 1/1  to: AL529152  from: 1  to: 844

      65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 AAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCT 50

      81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluT 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGA 100

      98 hrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThr 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACC 150

     115 GlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGl 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGA 200

     131 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValL 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGC 250

     148 euLeuAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeu 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 TGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTG 300

     165 CysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerTh 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTAC 350

     181 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGA 400

     198 leGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 214
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 TTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCA 450

     215 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySe 231
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATC 500

     231 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleG 248
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTATCG 550

     248 luThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGlu 264
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 AAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAG 600

     265 LysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVa 281
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 AAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 650

     281 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspL 298
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACC 700

     298 euAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 314
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGC 750

     315 ArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         |||||||||||||||||||||||||||||||||||||
     751 AGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCT 789

seq_name: gb_est2:BI822844

seq_documentation_block:
LOCUS       BI822844                928 bp    mRNA    linear   EST 04-OCT-2001
```

```
DEFINITION  603040166F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5180897 5',
            mRNA sequence.
ACCESSION   BI822844
VERSION     BI822844.1  GI:15934394
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11451  row: d  column: 18
            High quality sequence stop: 822.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5180897"
                     /clone_lib="NIH_MGC_115"
                     /lab_host="DH10B"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
BASE COUNT      198 a    256 c    259 g    214 t      1 others
ORIGIN


alignment_scores:
            Quality: 1357.50             Length:    295
              Ratio:  4.848               Gaps:      4
 Percent Similarity:  94.915   Percent Identity: 93.559

alignment_block:
 US-09-357-675C-21 x BI822844   ..

 Align seg 1/1  to: BI822844  from: 1  to: 928

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      39 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 88

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      89 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 138

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     139 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 188

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     189 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 238

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     239 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 288

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     289 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 338

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     339 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 388

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     389 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 438

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     439 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 488

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     489 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 538

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||   ||||||||||||||||||||||||||||||||||
     539 AGAGATTCCAGGCAG.GGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 587

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     588 GGCCCAGTCTTGAGTCACCTGTCAGCCCACCAGCAGGCAAGATTGGTCTA 637

     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     638 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 687

     217 nAlaGlyAlaGluIleLeuThr.TyrProSerAlaPheGlySer...Ile 232
         ||||||||||||||||||||||| ||||||||||||||||||||   |||
     688 AGCTGGAGCAGAGATACTTACCTTATCCTTCAGCTTTTGGATCCCATATA 737

     233 ThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     738 CAGGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGNCCGTGCTATCGAAAC 787

     249 rGlnCysTyrVal.ValAlaAlaAlaGlnCysGly.ArgHisHisGluLy 265
         |||||||||||| ||||||   |||||||||||| ||||||||||||||
     788 CCAGTGCTATGTTAGTGGCGCCAGCACAGTGTGGAACGCCACCATGAGAA 837

     265 sArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValV 282
         ||||   ||||||||||||||||||||||:::|||   |||      :
     838 GAGAGC.AGTTATGGGCACAGCATGGTGGTAAACCCTGGGGGACAGGGGG 886

     282 alAlaArgCysSerGluGlyProGlyLeu 291
         ::              ||||||||||||
     887 CCCGTGCTCTGAGGGCAGGCCCTGGCTTG 915

seq_name: gb_est2:BI769604

seq_documentation_block:
LOCUS       BI769604                 846 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603054985F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
            mRNA sequence.
ACCESSION   BI769604
VERSION     BI769604.1  GI:15761182
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
           Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11512  row: j  column: 22
           High quality sequence stop: 838.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5204469"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      179 a    240 c    226 g    201 t
ORIGIN


alignment_scores:
            Quality: 1308.00             Length:    264
              Ratio:  4.992               Gaps:      2
 Percent Similarity:  99.242   Percent Identity: 99.242

alignment_block:
 US-09-357-675C-21 x BI769604   ..

 Align seg 1/1  to: BI769604  from: 1  to: 846

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      59 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 108

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     109 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 158

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 208

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 258

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     259 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 308

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     309 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 358

     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     359 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 408

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     409 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 458

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     459 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 508

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     509 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 558

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     559 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 608

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     609 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 658

     201 AlaVal.CysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
         |||||| ||||||||||||||||||||||||||||||||||||||||||
     659 GCTGTCATGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTC 708

     217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     709 AAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 758

     234 GlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGl 250
         ||||||   ||||||||||||||||||||   ||||||||||||||||||
     759 GGCCCAGC.CACTGGGAGGTGTTGCTGCGGGC.CGTGCTATCGAAACCCA 806

     250 nCysTyrValValAlaAla.AlaGlnCysGlyArgHis 262
         |||||||||||||||||| ||||||||||||||||||
     807 GTGCTATGTAGTGGCAGCCAGCACAGTGTGGACGCCAC 844

seq_name: gb_est2:BG436916

seq_documentation_block:
LOCUS       BG436916                793 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602488424F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4620892 5',
            mRNA sequence.
ACCESSION   BG436916
VERSION     BG436916.1  GI:13343422
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 793)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1382  row: o  column: 05
            High quality sequence stop: 767.
FEATURES             Location/Qualifiers
     source          1. .793
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4620892"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
```

```
                  Superscript II RT (Life Technologies). Note: this is a
                  NIH_MGC Library."
BASE COUNT      181 a    210 c    230 g    172 t
ORIGIN


alignment_scores:
            Quality: 1298.00             Length:    267
              Ratio:  4.992               Gaps:      5
 Percent Similarity:  97.378   Percent Identity: 95.880

alignment_block:
 US-09-357-675C-21 x BG436916   ..

 Align seg 1/1  to: BG436916  from: 1  to: 793

      46 LeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnGl 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACA 52

      62 nAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyA 79
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTG 102

      79 laCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPro 95
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCT 152

      96 AlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGl 112
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGA 202

     112 uTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyP 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     203 ATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTT 252

     129 heHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCys 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGT 302

     146 HisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArgLysTh 162
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGAC 352

     162 rHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerA 179
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     353 ACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCA 402

     179 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 195
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     403 ACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCA 452

     196 GlyLysIleGlyLeuAlaValCysTyrAspMetArgProProGluLeuSe 212
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     453 GGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTC 502

     212 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro.SerAla 228
         |||||||||||||||||||||||||||||||||||||||||||| :::|||
     503 TCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTACAGCT 552

     229 PheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla.A 245
         |||||||||||||||||||||||||||||||||||||||||||||||| |
     553 TTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCC 602

     245 rgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArg 261
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     603 GTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGC 652

     262 HisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp.ProT 278
         |||||||||||||||||||||||||||||||||||||||||||| ||||
     653 CACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCCT 702

     278 rpGly.ThrValValAlaArgCysSer.GluGlyProGlyLeuCysLeuA 294
         |||||| |||||||||||||||||||| |||||||||||   :::|||
     703 GGGGAAACAGTGGTGGCCCGCTGCTCTTGAGGGGCCAGGCTCTGCCTTGC 752

     294 laArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
         |||||||||||:::|||||||||||||||||||||:::
     753 CCCGAATAGACCTCAAGTATCTGCGACAGTTGCGCCGAAAC 793

seq_name: gb_est1:AL559163

seq_documentation_block:
LOCUS       AL559163                861 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL559163 LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ014YA17 5
            prime, mRNA sequence.
ACCESSION   AL559163
VERSION     AL559163.1  GI:12904391
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ014YA17"
                     /clone_lib="LTI_NFL008_TC2"
                     /sex="male"
                     /tissue_type="T cells from T cell leukemia"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime  end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      176 a    229 c    225 g    201 t     30 others
ORIGIN


alignment_scores:
            Quality: 1253.00             Length:    280
              Ratio:  4.838               Gaps:      4
 Percent Similarity:  92.500   Percent Identity: 89.286

alignment_block:
 US-09-357-675C-21 x AL559163   ..

 Align seg 1/1  to: AL559163  from: 1  to: 861

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         |||||||||   ||||||||||||||||||||||||||||||||||||||
      28 ATGCTGGGCGTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 77

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      78 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 127

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     128 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 177

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
```

```
     |||||||||||||||||||||||||||||||||   ||||||||||||||
178 GTGTGCCAGGTAACATCGACGCCAGAYAAGCAAYAGAACTTTAAAACATG 227

 67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
228 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 277

 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
     ||||||||||||||||||||||||||   ||||||||||||||||||||
278 TGCCTGAGGCATTTGACTTCATTGCATGGGACCCTGCAGAGACGCTACAY 327

101 LeuSerGluPro.LeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
     ||||||||||| ||||||||||||||||||||||||||:::||||||:
328 CTGTCTGAACCTACTGGGTGGGAAACTTTTGGAAGAATACAYCCAGCTTK 377

117 laArgGlu.CysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGl 133
     ::|||::: |||||||||   |||   ||||||||||||||||||||||
378 CCAGGGATATGTGGACTCTKGCTGTTCTTGGGTGGTTTCCATGAGCGTGG 427

133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
     |   ||||||||||||||   |||||||||||||||||||||||||||||
428 CYAWGACTGGGAGCAGACTYAGAAAATCTACAATTGTCACGTGCTGCTGA 477

150 snSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAsp 166
     ||||||||||||||||||||||||||||||||||||||||||||||||||
478 ACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGAC 527

167 ValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183
     ||||||||||||||||||||||||||||||||||||||||||||||||||
528 GTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCC 577

183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
578 TGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTC 627

200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
     ||||||||||||||||||||||||||   |||||||||:::||||||
628 TAGCTGTCTGYTATGACATGCGGTTCCCTGCACTCTCTCTCRCATTGGCT 677

217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle.T 233
     ||||||||||||   |||:::      |||:::||||||:::||| |
678 CMAGCTGGCGCAGAGAWACTTWACTMWYCTTCMRCTTTTGGCWCCATTTA 727

233 hrGlyProAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluTh 249
     ||   |||||||||||| :::||||||||||||||||||||      ::
728 CARGCCCAGCCCACTGGGCAGGTGTTGCTGCGGGCCCGTGCTCWCGMAAM 777

249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
     :||||||   ||||||||||||||||||||||||||||||   |||||||
778 CCAGTGCTMWGTAGTKGCWGCAGCACAGTGTGGACGCCACMATGAGAAGA 827

266 rgAlaSerTyrGlyHisSerMetValValAsp 276
     |||||||||||||||||||||||||||||||
828 GAGCAAGTTATGGCCACAGCATGGTGGTAGAC 859

seq_name: gb_est2:BG762506

seq_documentation_block:
LOCUS       BG762506                764 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602733954F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4859292 5',
            mRNA sequence.
ACCESSION   BG762506
VERSION     BG762506.1  GI:14073159
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1714  row: l  column: 13
            High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4859292"
                     /clone_lib="NIH_MGC_49"
                     /tissue_type="melanotic melanoma, high MDR (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using  ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library. |"
BASE COUNT      162 a    218 c    201 g    183 t
ORIGIN


alignment_scores:
            Quality: 1218.00             Length:    240
              Ratio:  5.183                Gaps:      2
 Percent Similarity:  97.917   Percent Identity: 97.917

alignment_block:
 US-09-357-675C-21 x BG762506   ..

 Align seg 1/1  to: BG762506  from: 1  to: 764

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      49 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 98

      17 sProGlyLeuArgIleProGln.LeuSerValLeuCysAlaGlnProArg 33
         ||||||||||||||||||||    ||||||||||||||||||||||||||
      99 TCCTGGACTCCGGATACCTCTAACTCTCAGTACTTTGTGCTCAGCCCAGG 148

      34 ProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAl 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     149 CCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGC 198

      50 aValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrC 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     199 TGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACAT 248

      67 ysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     249 GTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTC 298

      84 LeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHi 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     299 CTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACA 348

     100 sLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     349 CCTGTCTGAACCACTGGGTGGGAAACTTTGGGAAGAATACACCCAGCTTG 398

     117 laArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     399 CCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGC 448
```

```
     134 GlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     449 CAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAA 498

     150 nSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspV 167
         |||||||||||||||||||||||||||||||||||||||||||||||||
     499 CAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACG 548

     167 alGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 183
         ||||||||||||   |||||||||||||||||||||||||||||||||||
     549 TAGAGATTCCACGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCT 598

     184 GlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     599 GGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCT 648

     200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     649 AGCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTC 698

     217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     699 AAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACA 748

     234 GlyPro.AlaHisTrp 238
         |||||| |||||||||
     749 GGCCCAGGCCCACTGG 764

seq_name: gb_est2:BG703176

seq_documentation_block:
LOCUS       BG703176                 790 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602686082F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4818405 5',
            mRNA sequence.
ACCESSION   BG703176
VERSION     BG703176.1  GI:13975252
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10720  row: d  column: 22
            High quality sequence stop: 766.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4818405"
                     /clone_lib="NIH_MGC_95"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT      167 a    228 c    206 g    189 t
ORIGIN


alignment_scores:
            Quality: 1206.00             Length:    238
              Ratio:  5.132                Gaps:      2
 Percent Similarity:  98.739   Percent Identity: 97.899

alignment_block:
 US-09-357-675C-21 x BG703176   ..

 Align seg 1/1  to: BG703176  from:`1  to: 790

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      79 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 128

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     129 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 178

      34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     179 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 228

      51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     229 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 278

      67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     279 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 328

      84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     329 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 378

     101 LeuSerGluProLeuGlyGlyLysLeuGluGluTyrThrGlnLeuAl 117
            ||||||||||||||||||||||||||||||||||||||||||||||
     379 CGGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 428

     117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
         ||||||||||||||||||||||||||||||||||||||:::||||||||
     429 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCATGGCC 478

     134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
         ||||||||||||||||||||||||||||||||||:::|||||||||||||
     479 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACATGCTGCTGAAC 528

     151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     529 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 578

     167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     579 AGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 628

     184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     629 GGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 678

     201 AlaValCysTyr.AspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
         |||||||||||| ||||||||||||||||||||||||||||||||||||
     679 GCTGTCTGCTATTGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTC 728

     217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle.Th 233
         |||||||||||||||||||||||||||||||||||||||||||||| ||
     729 AAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATCCATTTAC 778

     233 rGlyProAla 236
         |||||||||
```

```
     779 AGGCCCAGCC 788

seq_name: gb_est1:AL520767

seq_documentation_block:
LOCUS       AL520767                992 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 3
            prime, mRNA sequence.
ACCESSION   AL520767
VERSION     AL520767.1  GI:12784260
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      229 a    267 c    268 g    222 t      6 others
ORIGIN


alignment_scores:
            Quality: 1201.50            Length:    236
              Ratio:  5.179               Gaps:      3
 Percent Similarity:  98.305   Percent Identity: 96.610

alignment_block:
 US-09-357-675C-21 x AL520767/rev   ..

 Align seg 1/1  to reverse of: AL520767  from: 1  to: 992

      94 AspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLe 110
         ||||||||||||||||||||||||||||||||||||||||||:::   ::
     983 GACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGAAAC...TT 937

     110 uGluGlu.TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
         :   ::: ||||||||||||||||||||||||||||||||||||||||||
     936 TTGRAGRATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTG 887

     127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     886 GGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTA 837

     143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     836 CAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACA 787

     160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     786 GGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGT 737

     177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     736 GAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCAC 687

     193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     686 ACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTG 637

     210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     636 AACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCT 587

     227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     586 TCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCG 537

     243 gAlaArgAlaIleGluThrGlnCysTyrValVal.AlaAlaAlaGlnCys 259
         |||||||||||||||||||||||||||||||| :::|||||||||||||
     536 GGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGCCMCGCAGCACAGTGT 487

     260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAs 276
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     486 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGA 437

     276 pProTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysL 293
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     436 CCCCTGGGGAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCC 387

     293 euAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuPro 309
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     386 TTGCCCGAATAGACCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCT 337

     310 ValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLe 326
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     336 GTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACT 287

     326 uSer 327
         ||||
     286 GTCT 283

seq_name: gb_est1:AL522373

seq_documentation_block:
LOCUS       AL522373                960 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3
            prime, mRNA sequence.
ACCESSION   AL522373
VERSION     AL522373.1  GI:12785866
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB008YM02"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
```

```
                /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector.  Library was normalized. Library was constructed
                by Life Technologies.  Contact : Feng Liang Life
                Technologies, a division of Invitrogen 9800 Medical Center
                Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                8371 Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT      225 a    259 c    255 g    217 t      4 others
ORIGIN


alignment_scores:
           Quality: 1179.50             Length:    231
             Ratio:  5.151                Gaps:      2
 Percent Similarity:  99.134   Percent Identity: 98.268

alignment_block:
 US-09-357-675C-21 x AL522373/rev   ..

 Align seg 1/1  to reverse of: AL522373  from: 1  to: 960

      98 ThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeu.GluGluTyrT 114
         ||||||||||||||||||||||||||||||:::   ::: ||||||||||
     959 ACGCTACACCTGTCTGAACCACTGGGTGGAAAC...TTTTGAAGAATACA 913

     114 hrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHis 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     912 CCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCAT 863

     131 GluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVa 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     862 GAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGT 813

     147 lLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisL 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     812 GCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATC 763

     164 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     762 TGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCT 713

     181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     712 ACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAA 663

     197 sIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuA 214
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     662 GATTGGTCTAGCTGTCTGCTATGACATGCG.TTCCCTGAACTCTCTCTGG 614

     214 laLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 230
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     613 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGA 564

     231 SerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIl 247
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     563 TCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTGCTAT 514

     247 eGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisG 264
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 CGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGCCACCATG 464

     264 luLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     463 AGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACA 414

     281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAs 297
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     413 GTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGA 364

     297 pLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisA 314
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     363 CCTCAACTATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACC 314

     314 rgArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
         ||||||||||||||||||||||||||||||||||||||||
     313 GCAGGCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCT 273

seq_name: gb_est2:BI757823

seq_documentation_block:
LOCUS       BI757823                946 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603030404F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200560 5',
            mRNA sequence.
ACCESSION   BI757823
VERSION     BI757823.1  GI:15749401
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 946)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11502  row: h  column: 01
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .946
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5200560"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      213 a    264 c    270 g    199 t
ORIGIN


alignment_scores:
           Quality: 1169.00             Length:    291
             Ratio:  4.395                Gaps:      7
 Percent Similarity:  91.409   Percent Identity: 83.849

alignment_block:
 US-09-357-675C-21 x BI757823   ..

 Align seg 1/1  to: BI757823  from: 1  to: 946

       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      69 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 118

      17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     119 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 168
```

```
     34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    169 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 218

     51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    219 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 268

     67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    269 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 318

     84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    319 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 368

    101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    369 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 418

    117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    419 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 468

    134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
        ||||||||||||||||||||||||||||:::|||||||||||||||||||
    469 AAGACTGGGAGCAGACTCAGAAAATCTACAAGTGTCACGTGCTGCTGAAC 518

    151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    519 AGCAAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 568

    167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
        ||||||||||||   ||||||||||||||||||||||||||||||||||
    569 AGAGATTCCAGGCAG.GGGCCTATGTGTGAAAGCAACTCTACCATGCCTG 617

    184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    618 GGCCCAGTCTTGAGTCACCTGTCAGCACACCGGCAGGCAAGATTGGTCTA 667

    201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
        ||||||||||||||||||||||||||:::      ::: ' :::||
    668 GCTGTCTGCTATGACATGCGGTTCCCTGAAATTCTCTCGGGCAGTGGTCA 717

    217 nAlaGlyAlaGluIleLeuThr.TyrProSerAlaPheGly.SerIleTh 233
        |||||||||||||||||||| ||||||||||||:::    ::::::||
    718 AGCTGGAGCAGAGATACTTACCGTATCCGTCAGCTTGGGATCCAGTTAAC 767

    233 rGlyPro.........AlaHisTrpGluValLeuLeuArgAlaArgAlaI 247
        |||||||         ::::::|||:::      ||||||||::::::
    768 GGGGCCCAGACCACTGGGAAGGTGGAAG......CTGAGGGCCCAGTGCC 811

    247 le...GluThr...GlnCysTyrValValAlaAlaAlaGlnCysGlyArg 261
        ::   ||||||   ::::::   :::|||:::   |||:::   ||||||
    812 TATCCGAAACGCTAAGGGCTCAGATAGTGGGCAGAGCACACGG.GGACGC 860

    262 His...HisGluLysArgAlaSerTyrGlyHisSerMetValValAspPr 277
        |||      ||||||      :::   ||||||||||||||      ||
    861 CACCCTGACGAGAAGAGCCAAGATAAGGGCCACAGCATGGTGGAAAAACC 910

    277 oTrpGlyThrValValAla 283
        |   |||:::   :::|||
    911 CATGGGAAACAGGATGGCC 929

seq_name: gb_est2:BI755493

seq_documentation_block:
LOCUS       BI755493                813 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603027349F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5197716 5',
            mRNA sequence.
ACCESSION   BI755493
VERSION     BI755493.1  GI:15747071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11495  row: a  column: 13
           High quality sequence stop: 745.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5197716"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      155 a    241 c    225 g    192 t
ORIGIN


alignment_scores:
            Quality: 1135.50            Length:    243
              Ratio:  4.916              Gaps:      2
 Percent Similarity:  95.062   Percent Identity: 92.593

alignment_block:
 US-09-357-675C-21 x BI755493   ..

 Align seg 1/1  to: BI755493  from: 1  to: 813

      1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     87 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTG 136

     17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    137 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGC 186

     34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    187 CCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCCCTGGTGGCT 236

     51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    237 GTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATG 286

     67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    287 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCC 336

     84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    337 TGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 386
```

```
    101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    387 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGC 436

    117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    437 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCC 486

    134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    487 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCACGTGCTGCTGAAC 536

    151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
        |||   ||||||||||||||||||||||||||||||||||||||||||||
    537 AGCACCGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 586

    167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
        ||||||||||||||||||||||||||||   ||||||||||||||||||||
    587 AGAGATTCCAGGGCAGGGGCCTATGTGTGCAAGCAACTCTACCATGCCTG 636

    184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
        ||||||||||||||||||||||||||:::   ||||||   ||||||||||
    637 GGCCCAGTCTTGAGTCACCTGTCAGCGCAGCAGCAGGCAG.ATTGGGCTA 685

    201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
        ||||||||||||||:::||||||||||||||||||||||||||||||||
    686 GCTGTCTGCTATGACGTGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCA 735

    217 n..AlaGlyAlaGluIleLeuThrTyrProSerAlaPhe..........  229
        |  ||||||||||||||||||||||||||||||:::
    736 GACGCTGGAGCCGAGATACTTACCTATCCTTCAGCTTTGCGTCCTTGCGG 785

    230 ..........GlySerIleThrGlyPro 235
                |||::::::   ||||||
    786 GCCCGCCACTGGAGCGGTGCTGGGGCCT 813

seq_name: gb_est2:BI711300

seq_documentation_block:
LOCUS       BI711300                625 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  id96f12.x1 Human insulinoma Homo sapiens cDNA 3' similar to
            TR:O76091 O76091 NITRILASE HOMOLOG 1. ;, mRNA sequence.
ACCESSION   BI711300
VERSION     BI711300.1  GI:15686995
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human insulinoma"
                     /tissue_type="insulinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
                     (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
                     pBluescript SK- by Dr. H. Inoue following the Washington
                     University protocol
                     (http://genome.wustl.edu/est/lambda_protocol.shtml).
                     Please contact Hiroshi Inoue, MD/PhD for further
                     information on this library (Metabolism Division, Permutt
                     Laboratory, Washington University School of Medicine, Box
                     8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
                     is a Washington University Pancreas EST project library. "
BASE COUNT      139 a    183 c    161 g    142 t
ORIGIN


alignment_scores:
           Quality: 1115.00            Length:    208
             Ratio:  5.361              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-357-675C-21 x BI711300/rev   ..

 Align seg 1/1  to reverse of: BI711300  from: 1  to: 625

     76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAl 92
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    625 AGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGC 576

     92 aArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysL 109
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    575 ACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAAAC 526

    109 euLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSer 125
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    525 TTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCC 476

    126 LeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIl 142
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    475 TTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAAAT 426

    142 eTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrT 159
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    425 CTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTT 376

    159 yrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    375 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 326

    176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    325 TGTGAAAGCAACTCTACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAG 276

    192 rThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheP 209
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    275 CACACCAGCAGGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCC 226

    209 roGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyr 225
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    225 CTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTAT 176

    226 ProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLe 242
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    175 CCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTGCT 126

    242 uArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnC 259
        |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
125 GCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGT 76

259 ysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValVal 275
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 75 GTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTA 26

276 AspProTrpGlyThrValValAla 283
    ||||||||||||||||||||||||
 25 GACCCCTGGGGAACAGTGGTGGCC 2
```

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2002, 10:31:12 ; Search time 2667 Seconds
(without alignments)
3568.286 Million cell updates/sec

Title: US-09-357-675C-21
Perfect score: 327
Sequence: 1 MLGFITRPPHRFLSLLCPGL.....LPVFQHRRPDLYGNLGHPLS 327

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09357675/runat_03122002_142235_20398/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09357675_@CGN_1_1_1616_@runat_03122002_142235_20398 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		327	100.0	1385	9	AF069987	AF069987 Homo sapi
2		88	26.9	4079	9	AF069984	AF069984 Homo sapi
3	c	88	26.9	167863	2	AL590651	AL590651 Homo sapi
4		88	26.9	200822	9	AL591806	AL591806 Human DNA
5		66	20.2	847	11	G72919	G72919 MARC 2849-2
6	c	63	19.3	377	6	AX368386	AX368386 Sequence
7		33	10.1	1338	10	AF069988	AF069988 Mus muscu
8		33	10.1	1365	10	BC021634	BC021634 Mus muscu
9		33	10.1	4481	10	AF069985	AF069985 Mus muscu
10		33	10.1	211772	10	AC084821	AC084821 Mus muscu
11	c	33	10.1	215043	10	AC087229	AC087229 Mus muscu
12		23	7.0	1214	5	AF284575	AF284575 Xenopus l
13		23	7.0	1214	6	AX419587	AX419587 Sequence
14	c	18	5.5	115667	2	AC125857	AC125857 Rattus no
15		18	5.5	181583	2	AC105589	AC105589 Rattus no
16		16	4.9	10881	1	AE004129	AE004129 Vibrio ch
17		13	4.0	924	8	AY133544	AY133544 Arabidops
18		13	4.0	1172	8	AF372904	AF372904 Arabidops
19	c	13	4.0	104386	8	ATT32A17	AL161813 Arabidops
20	c	13	4.0	126375	2	CNS08C81	AL731752 Oryza sat
21		13	4.0	137775	2	CNS08C80	AL731751 Oryza sat
22	c	13	4.0	199362	8	ATCHRIV24	AL161512 Arabidops
23		12	3.7	1385	3	AF069986	AF069986 Caenorhab
24		12	3.7	169364	3	CEY56A3A	AL132860 Caenorhab
25		11	3.4	1521	3	AF069989	AF069989 Drosophil
26		11	3.4	1576	3	AY089221	AY089221 Drosophil
27		11	3.4	25350	8	SPBC651	AL035570 S.pombe c
28	c	11	3.4	34013	2	AC010013	AC010013 Drosophil
29		11	3.4	131853	2	AC017755	AC017755 Drosophil
30	c	11	3.4	168250	3	AC093121	AC093121 Drosophil
31		11	3.4	173497	3	AC093501	AC093501 Drosophil
32		11	3.4	174279	2	AC006169	AC006169 Drosophil
33	c	11	3.4	186793	2	AC095657	AC095657 Rattus no
34		11	3.4	190353	3	AC105293	AC105293 Drosophil
35		11	3.4	298640	3	AE003467	AE003467 Drosophil
36	c	10	3.1	10466	1	AE004861	AE004861 Pseudomon
37	c	10	3.1	15124	1	AE000934	AE000934 Methanoba
38	c	10	3.1	142756	9	AC005844	AC005844 Homo sapi
39		10	3.1	149000	2	AC129010	AC129010 Leishmani
40		9	2.8	1520	8	AF284572	AF284572 Saccharom
41		9	2.8	4123	1	EAC271724	AJ271724 Eubacteri
42	c	9	2.8	5310	9	HSM803919	AL832609 Homo sapi
43	c	9	2.8	6597	6	E49884	E49884 Analysis me
44	c	9	2.8	6597	9	HSCAN	X64228 H.sapiens c
45	c	9	2.8	6642	9	HUMORF10	D14689 Human mRNA

ALIGNMENTS

RESULT 1

```
AF069987
LOCUS       AF069987               1385 bp    mRNA    linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.
ACCESSION   AF069987
VERSION     AF069987.1  GI:3228665
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1385)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            1. .1385
                     /gene="NIT1"
     CDS             77. .1060
                     /gene="NIT1"
                     /note="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC39907.1"
                     /db_xref="GI:3228666"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
BASE COUNT      322 a    382 c    357 g    324 t
ORIGIN

Alignment Scores:
Pred. No.:              0           Length:         1385
Score:                  327.00      Matches:        327
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
Query Match:            100.00%     Indels:         0
DB:                     9           Gaps:           0

US-09-357-675C-21 (1-327) x AF069987 (1-1385)

Qy       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeu 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTC 136

Qy      21 ArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 CGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCC 196

Qy      41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLys 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 TCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAG 256

Qy      61 GlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CAACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGC 316

Qy      81 LeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

Qy     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCys 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGT 436

Qy     121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 GGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAG 496

Qy     141 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 AAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGG 556

Qy     161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 AAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCT 616

Qy     181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     617 ACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 676

Qy     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     677 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCA 736

Qy     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     737 GAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTG 796

Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGly 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     797 TTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGA 856

Qy     261 ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     857 CGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACA 916

Qy     281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     917 GTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 976

Qy     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     977 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGC 1036

Qy     321 AsnLeuGlyHisProLeuSer 327
           |||||||||||||||||||||
Db    1037 AATCTGGGTCACCCACTGTCT 1057

RESULT 2
AF069984
LOCUS       AF069984               4079 bp    DNA     linear   PRI 23-JUL-1998
DEFINITION  Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069984
VERSION     AF069984.1  GI:3242977
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
```

```
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 4079)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21-q22"
     gene            635. .3569
                     /gene="NIT1"
     mRNA            join(635. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3569)
                     /gene="NIT1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            635. .680
                     /gene="NIT1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(679. .680,1288. .1383,1636. .1890,2011. .2114,2321. .2454,
                     2609. .2734,3001. .3267)
                     /gene="NIT1"
                     /note="Nit1; alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC39901.1"
                     /db_xref="GI:3242978"
                     /translation="MLGFITRPPHRFLSLLCPGLRIPQLSVLCAQPRPRAMAISSSSC
                     ELPLVAVCQVTSTPDKQQNFKTCAELVREAARLGACLAFLPEAFDFIARDPAETLHLS
                     EPLGGKLLEEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVATYR
                     KTHLCDVEIPGQGPMCESNSTMPGPSLESPVSTPAGKIGLAVCYDMRFPELSLALAQA
                     GAEILTYPSAFGSITGPAHWEVLLRARAIETQCYVVAAAQCGRHHEKRASYGHSMVVD
                     PWGTVVARCSEGPGLCLARIDLNYLRQLRRHLPVFQHRRPDLYGNLGHPLS"
     exon            681. .711
                     /gene="NIT1"
                     /note="1A; alternatively spliced"
     exon            712. .741
                     /gene="NIT1"
                     /note="1B; alternatively spliced"
     exon            940. .1053
                     /gene="NIT1"
                     /note="1C; alternatively spliced"
     exon            1288. .1383
                     /gene="NIT1"
                     /number=2
     exon            1636. .1890
                     /gene="NIT1"
                     /number=3
     exon            2011. .2114
                     /gene="NIT1"
                     /number=4
     exon            2321. .2454
                     /gene="NIT1"
                     /number=5
     exon            2609. .2734
                     /gene="NIT1"
                     /number=6
     exon            3001. .3569
                     /gene="NIT1"
                     /number=7
BASE COUNT      940 a   1082 c   1043 g   1012 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              6.87e-85        Length:         4079
```

```
Score:                  88.00           Matches:        88
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            26.91%          Indels:         0
DB:                     9               Gaps:           0

US-09-357-675C-21 (1-327) x AF069984 (1-4079)

Qy     240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 3060

Qy     260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 3120

Qy     280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 3180

Qy     300 TyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyr 319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 3240

Qy     320 GlyAsnLeuGlyHisProLeuSer 327
           |||||||||||||||||||||||||
Db    3241 GGCAATCTGGGTCACCCACTGTCT 3264

RESULT 3
AL590651/c
LOCUS       AL590651              167863 bp    DNA     linear   HTG 18-AUG-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL590651
VERSION     AL590651.4  GI:13990236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167863)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13751001.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA137A12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 164729 bases at least Q40
            Consensus quality: 165486 bases at least Q30
            Consensus quality: 166169 bases at least Q20
            Insert size: 167063; sum-of-contigs
            Insert size: 171878; 3.4% error; agarose-fp
            Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
            coverage: 8.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
           * be preserved.
           *        1     9373: contig of 9373 bp in length
           *     9374 9473: gap of      100 bp
           *     9474    29891: contig of 20418 bp in length
           *    29892 29991: gap of      100 bp
           *    29992    36239: contig of 6248 bp in length
           *    36240 36339: gap of      100 bp
           *    36340   103784: contig of 67445 bp in length
           *   103785 103884: gap of      100 bp
           *   103885   114878: contig of 10994 bp in length
           *   114879 114978: gap of      100 bp
           *   114979   126347: contig of 11369 bp in length
           *   126348 126447: gap of      100 bp
           *   126448   137764: contig of 11317 bp in length
           *   137765 137864: gap of      100 bp
           *   137865   163160: contig of 25296 bp in length
           *   163161 163260: gap of      100 bp
           *   163261   167863: contig of 4603 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-137A12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .9373
                     /note="assembly_fragment:02360
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9474. .29891
                     /note="assembly_fragment:01854
                     fragment_chain:1"
     misc_feature    29992. .36239
                     /note="assembly_fragment:00805
                     fragment_chain:1"
     misc_feature    36340. .103784
                     /note="assembly_fragment:01467
                     fragment_chain:1"
     misc_feature    103885. .114878
                     /note="assembly_fragment:00991
                     fragment_chain:2"
     misc_feature    114979. .126347
                     /note="assembly_fragment:01503
                     fragment_chain:2"
     misc_feature    126448. .137764
                     /note="assembly_fragment:01297
                     fragment_chain:2"
     misc_feature    137865. .163160
                     /note="assembly_fragment:00117
                     fragment_chain:2"
     misc_feature    163261. .167863
                     /note="assembly_fragment:01267
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    43927 a  39935 c  40002 g  43197 t    802 others
ORIGIN

Alignment Scores:
Pred. No.:              1.98e-83        Length:         167863
Score:                  88.00           Matches:        88
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            26.91%          Indels:         0
DB:                     2               Gaps:           0

US-09-357-675C-21 (1-327) x AL590651 (1-167863)

Qy    240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149680 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 149621

Qy    260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149620 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 149561

Qy    280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149560 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 149501

Qy    300 TyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyr 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149500 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 149441

Qy    320 GlyAsnLeuGlyHisProLeuSer 327
          |||||||||||||||||||||||||
Db 149440 GGCAATCTGGGTCACCCACTGTCT 149417

RESULT 4
AL591806
LOCUS       AL591806               200822 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
            sequence.
ACCESSION   AL591806
VERSION     AL591806.16  GI:18476709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200822)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17902927.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-544M22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-544M22 The true
            left end of clone RP11-137A12 is at 156538 in this sequence. The
            true right end of clone RP11-381D2 is at 145015 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .200822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-544M22"
                     /clone_lib="RPCI-11.2"
     misc_feature    10395. .10397
                     /note="Sequence from overlapping clone RP11-381D2
```

```
                (AL162592). Assembly confirmed by restriction digest."
     misc_feature    44098. .44107
                /note="Sequence from overlapping clone RP11-381D2
                (AL162592). Assembly confirmed by restriction digest."
     misc_feature    59585. .59604
                /note="Sequence from overlapping clone RP11-381D2
                (AL162592). Assembly confirmed by restriction digest."
     misc_feature    90797. .90883
                /note="Single clone region. Reads generated from a
                transposon library derived from a single pUC clone.
                Restriction digest data confirm the assembly."
     misc_feature    90861. .90883
                /note="Sequence from uni-directional dGTP big dye
                terminator reads only."
BASE COUNT    55215 a  49052 c  46183 g  50372 t
ORIGIN

Alignment Scores:
Pred. No.:              2.33e-83        Length:         200822
Score:                  88.00           Matches:        88
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            26.91%          Indels:         0
DB:                     9               Gaps:           0

US-09-357-675C-21 (1-327) x AL591806 (1-200822)

Qy     240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174501 GTGTTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGT 174560

Qy     260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174561 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGA 174620

Qy     280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174621 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAAC 174680

Qy     300 TyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyr 319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174681 TATCTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTAT 174740

Qy     320 GlyAsnLeuGlyHisProLeuSer 327
           |||||||||||||||||||||||||
Db 174741 GGCAATCTGGGTCACCCACTGTCT 174764

RESULT 5
G72919
LOCUS       G72919                   847 bp    DNA     linear   STS 08-AUG-2001
DEFINITION  MARC 2849-2850:991933517:1 SCF - porcine spleen Sus scrofa STS
            genomic, sequence tagged site.
ACCESSION   G72919
VERSION     G72919.1  GI:15146949
KEYWORDS    STS.
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 847)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
            Keele,J.W.
  TITLE     Single nucleotide polymorphism (SNP) discovery in expressed porcine
            genes
  JOURNAL   Unpublished
COMMENT
            Contact: Freking BA
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4278
            Fax: 402 762 4173
            Email: freking@email.marc.usda.gov
            Primer A: GGGTTCAGCTTTTGGATCTG
            Primer B: GCTGAGCCTGTGTTTCATCA
            STS size: 700
            PCR Profile:
                 Hotstart:  95 degrees for 15 minutes
                 Denature:  95 degrees for 30 seconds
                 Anneal:    58 degrees
                 Extension: 68 degrees for 2 minutes
                 Cycles:    32 to 45
            Protocol:
                 Template:       50-200 ng genomic DNA
                 Primer:         each 20 pmoles
                 dNTPs:          each 88 uM
                 Taq Polymerase: 0.25 units (Qiagen HotStar)

            Buffer:
                 Commercially supplied Qiagen HotStar buffer

            The STS is derived from PCR amplicons generated from genomic DNA,
            sequenced from each end using the amplification primers.  The
            sequence does not necessarily represent the entire amplicon.
            Sequence derived from PolyPhred was trimmed from each end of each
            unique contig until five consecutive bases exceeded a quality score
            threshold of 20, and the next 10 bases averaged a quality score of
            20 or greater. Amplicon size was estimated by agarose gel
            electrophoresis.
FEATURES             Location/Qualifiers
     source          1. .847
                     /organism="Sus scrofa"
                     /strain="white composite, duroc, meishan, minzhu,
                     fengjing, crossbreds"
                     /db_xref="taxon:9823"
                     /sex="male and female"
                     /clone_lib="SCF - porcine spleen"
                     /dev_stage="adult"
                     /note="Organ: spleen"
     STS             <1. .>847
BASE COUNT      205 a    249 c    200 g    193 t
ORIGIN

Alignment Scores:
Pred. No.:              1.3e-61         Length:         847
Score:                  66.00           Matches:        66
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            20.18%          Indels:         0
DB:                     11              Gaps:           0

US-09-357-675C-21 (1-327) x G72919 (1-847)

Qy     240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 GTGTTGCTGCGGGCCCGAGCCATTGAAACTCAGTGCTATGTAGTGGCGGCAGCACAGTGT 543

Qy     260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 GGACGCCACCATGAGAAGAGAGCAAGTTATGGCCATAGCATGGTGGTAGATCCCTGGGGA 603

Qy     280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     604 ACGGTGGTGGCCCGCTGCTCTGAAGGACCAGGCCTCTGCCTTGCCCGAATTGACCTCAAT 663

Qy     300 TyrLeuArgGlnLeuArg 305
           |||||||||||||||||
Db     664 TATCTGCGACAGTTGCGC 681

RESULT 6
AX368386/c
LOCUS       AX368386                 377 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1096 from Patent WO0204514.
ACCESSION   AX368386
VERSION     AX368386.1  GI:18856459
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1096 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       94 a    100 c    112 g     71 t
ORIGIN

Alignment Scores:
Pred. No.:              1.13e-58        Length:         377
Score:                  63.00           Matches:        124
Percent Similarity:     98.41%          Conservative:   0
Best Local Similarity:  98.41%          Mismatches:     1
Query Match:            19.27%          Indels:         2
DB:                     6               Gaps:           0

US-09-357-675C-21 (1-327) x AX368386 (1-377)

Qy       8 ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 27
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy      28 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuPro 47
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy      48 LeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy      68 AlaGlu-LeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAl 87
           |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGG-GCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGC 139

Qy      87 aPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGl 107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGG 79

Qy     107 yLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGG 19

Qy     127 yGlyPheHisGluArg 132
           |||||||||||||||
Db      18 TGGTTTCCATGAGCGT 3

RESULT 7
AF069988
LOCUS       AF069988               1338 bp    mRNA    linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase 1 (Nit1) mRNA, complete cds.
ACCESSION   AF069988
VERSION     AF069988.1  GI:3228667
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 1338)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1338
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            1. .1338
                     /gene="Nit1"
     CDS             58. .1029
                     /gene="Nit1"
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAC40185.1"
                     /db_xref="GI:3228668"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      347 a    335 c    330 g    326 t
ORIGIN

Alignment Scores:
Pred. No.:              1.36e-25        Length:         1338
Score:                  33.00           Matches:        33
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            10.09%          Indels:         0
DB:                     10              Gaps:           0

US-09-357-675C-21 (1-327) x AF069988 (1-1338)

Qy     266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGC 900

Qy     286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
           |||||||||||||||||||||||||||||||||||||
Db     901 TCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTC 939

RESULT 8
BC021634
LOCUS       BC021634               1365 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
            complete cds.
ACCESSION   BC021634
VERSION     BC021634.1  GI:18204912
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6754855.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:13825 IMAGE:4008543"
                     /tissue_type="Mammary tumor metastatized to lung.
                     MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
                     enhancer."
                     /clone_lib="NCI_CGAP_Lu30"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             59. .1030
                     /codon_start=1
                     /product="nitrilase 1"
                     /protein_id="AAH21634.1"
                     /db_xref="GI:18204913"
                     /db_xref="LocusID:27045"
                     /translation="MLGFITRPPHQLLCTGYRLLRTPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYSSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
BASE COUNT      368 a    334 c    331 g    332 t
ORIGIN

Alignment Scores:
Pred. No.:              1.39e-25        Length:         1365
Score:                  33.00           Matches:        33
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            10.09%          Indels:         0
DB:                     10              Gaps:           0

US-09-357-675C-21 (1-327) x BC021634 (1-1365)

Qy     266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 AGAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGC 901

Qy     286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
           |||||||||||||||||||||||||||||||||||||
Db     902 TCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTC 940

RESULT 9
AF069985
LOCUS       AF069985               4481 bp    DNA     linear   ROD 23-JUL-1998
DEFINITION  Mus musculus nitrilase homolog 1 (Nit1) gene, alternatively spliced
            product, complete cds.
ACCESSION   AF069985
VERSION     AF069985.1  GI:3242979
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Nitrilase and Fhit homologs are encoded as fusion proteins in
            Drosophila melanogaster and Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
  MEDLINE   98337986
   PUBMED   9671749
REFERENCE   2  (bases 1 to 4481)
  AUTHORS   Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
            Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
            Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
            Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .4481
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1q21-q23"
     gene            541. .3944
                     /gene="Nit1"
     mRNA            join(541. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3944)
                     /gene="Nit1"
                     /product="nitrilase homolog 1"
                     /note="alternatively spliced"
     exon            541. .606
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=1
     CDS             join(605. .606,1177. .1263,1638. .1889,2015. .2118,2362. .2495,
                     2626. .2751,3392. .3658)
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="nitrilase homolog 1"
                     /protein_id="AAC40184.1"
                     /db_xref="GI:3242980"
                     /translation="MLGFITRPPHQLLCTGYRLLRIPVLCTQPRPRTMSSSTSWELPL
                     VAVCQVTSTPNKQENFKTCAELVQEAARLGACLAFLPEAFDFIARNPAETLLLSEPLN
                     GDLLGQYSQLARECGIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHL
                     CDVEIPGQGPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAQAGAEI
                     LTYPSAFGSVTGPAHWEVLLRARAIESQCYVIAAAQCGRHHETRASYGHSMVVDPWGT
                     VVARCSEGPGLCLARIDLHFLQQMRQHLPVFQHRRPDLYGSLGHPLS"
     exon            863. .986
                     /gene="Nit1"
                     /note="alternatively spliced"
                     /number=2
     exon            1177. .1263
                     /gene="Nit1"
                     /number=3
     exon            1638. .1889
                     /gene="Nit1"
                     /number=4
     exon            2015. .2118
                     /gene="Nit1"
                     /number=5
     exon            2362. .2495
                     /gene="Nit1"
                     /number=6
     exon            2626. .2751
                     /gene="Nit1"
                     /number=7
     exon            3392. .3944
                     /gene="Nit1"
                     /number=8
BASE COUNT     1101 a   1132 c   1133 g   1115 t
ORIGIN

Alignment Scores:
Pred. No.:              4.07e-25        Length:         4481
Score:                  33.00           Matches:        33
```

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.09%         Indels:        0
DB:                     10             Gaps:          0

US-09-357-675C-21 (1-327) x AF069985 (1-4481)

Qy    266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 AGAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGC 3529

Qy    286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
          |||||||||||||||||||||||||||||||||||||||
Db   3530 TCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTC 3568

RESULT 10
AC084821
LOCUS       AC084821              211772 bp    DNA     linear   ROD 19-FEB-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-395h6, complete sequence.
ACCESSION   AC084821
VERSION     AC084821.25  GI:18702388
KEYWORDS    HTG.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-395h6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 211772)
  AUTHORS   Deschamps,S., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 19, 2002 this sequence version replaced gi:18390260.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .211772
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-395h6"
                     /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT    55607 a  48743 c  49203 g  58215 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:              1.33e-23       Length:        211772
Score:                  33.00          Matches:       33
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.09%         Indels:        0
DB:                     10             Gaps:          0

US-09-357-675C-21 (1-327) x AC084821 (1-211772)

Qy    266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2809 AGAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGC 2868

Qy    286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
          |||||||||||||||||||||||||||||||||||||||
Db   2869 TCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTC 2907

RESULT 11
AC087229/c
LOCUS       AC087229              215043 bp    DNA     linear   ROD 14-AUG-2002
DEFINITION  Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,
            complete sequence.
ACCESSION   AC087229
VERSION     AC087229.24  GI:22094398
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Mus musculus Chromosome 1 BAC Clone rp23-191a19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 215043)
  AUTHORS   Jiang,X., Song,L., Gu,W. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Aug 3, 2002 this sequence version replaced gi:22038565.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .215043
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-191a19"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT    55806 a  49340 c  51162 g  58732 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              1.35e-23       Length:        215043
Score:                  33.00          Matches:       33
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            10.09%         Indels:        0
DB:                     10             Gaps:          0
```

```
US-09-357-675C-21 (1-327) x AC087229 (1-215043)

Qy    266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  88091 AGAGCAAGTTATGGCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGGTGGCCCGCTGC 88032

Qy    286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
          |||||||||||||||||||||||||||||||||||||
Db  88031 TCCGAGGGACCAGGCCTCTGCCTTGCTCGAATTGATCTC 87993

RESULT 12
AF284575
LOCUS       AF284575               1214 bp    mRNA    linear   VRT 22-JUL-2000
DEFINITION  Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION   AF284575
VERSION     AF284575.1  GI:9367117
KEYWORDS    .
SOURCE      Xenopus laevis.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1214)
  AUTHORS   Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
  TITLE     Crystal structure of the worm NitFhit Rosetta Stone protein reveals
            a Nit tetramer binding two Fhit dimers
  JOURNAL   Curr. Biol. 10 (15), 907-917 (2000)
  MEDLINE   20414396
   PUBMED   10959838
REFERENCE   2  (bases 1 to 1214)
  AUTHORS   Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson
            University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
     source          1. .1214
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1214
                     /gene="Nit1"
     CDS             180. .1046
                     /gene="Nit1"
                     /note="Nit1; similar to the Nit domains of Caenorhabditis
                     elegans NitFhit and Drosophila melanogaster NitFhit"
                     /codon_start=1
                     /product="Nit protein 1"
                     /protein_id="AAF87104.1"
                     /db_xref="GI:9367118"
                     /translation="MAGAHKPLIAVCQMTSTSDKEKNFATCSRLIREAAGRRACMVFL
                     PEAFDYIGGSIEETLSLAESLHGDTIQRYTQLARECGLWLSLGGFHEKGPNWDTDQRI
                     SNSHVVVDNTGHIVSVYRKAHLFDVDLQNGVSLRESSSTLPGAELIRPITSPAGKIGL
                     GVCYDLRFPEFSLALAQQGAELLTYPSAFTLTTGLAHWEVLLRARAIETQCYVVAAAQ
                     TDRHNEKRTSYGHAMVVDPWGLVIGQCQEGTGICYAEIDIPYMERVRRDMPVWRHRRT
                     DLYGKISFNKPD"
BASE COUNT      281 a    305 c    345 g    282 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              9.07e-15        Length:         1214
Score:                  23.00           Matches:        23
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            7.03%           Indels:         0
DB:                     5               Gaps:           0

US-09-357-675C-21 (1-327) x AF284575 (1-1214)

Qy    236 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAla 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    765 GCACATTGGGAGGTGTTGCTGAGAGCCCGTGCCATAGAAACCCAGTGCTACGTAGTTGCA 824

Qy    256 AlaAlaGln 258
          ||||||||
Db    825 GCGGCACAG 833

RESULT 13
AX419587
LOCUS       AX419587               1214 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 10 from Patent WO0187958.
ACCESSION   AX419587
VERSION     AX419587.1  GI:21523970
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1
  AUTHORS   Croce,C., Brenner,C. and Pekarski,Y.
  TITLE     Crystal structure of worm nitfhit reveals that a nit tetramer binds
            two fhit dimers
  JOURNAL   Patent: WO 0187958-A 10 22-NOV-2001;
            Thomas Jefferson University (US)
FEATURES             Location/Qualifiers
     source          1. .1214
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
BASE COUNT      281 a    305 c    345 g    282 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              9.07e-15        Length:         1214
Score:                  23.00           Matches:        23
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            7.03%           Indels:         0
DB:                     6               Gaps:           0

US-09-357-675C-21 (1-327) x AX419587 (1-1214)

Qy    236 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAla 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    765 GCACATTGGGAGGTGTTGCTGAGAGCCCGTGCCATAGAAACCCAGTGCTACGTAGTTGCA 824

Qy    256 AlaAlaGln 258
          ||||||||
Db    825 GCGGCACAG 833

RESULT 14
AC125857/c
LOCUS       AC125857             115667 bp    DNA     linear   HTG 23-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-421M5, *** SEQUENCING IN PROGRESS
            ***, 44 unordered pieces.
ACCESSION   AC125857
VERSION     AC125857.1  GI:21671452
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 115667)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
```

```
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 115667)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 115667)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWUD
                Center clone name: CH230-421M5
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 72804 bases at least Q40
                Consensus quality: 77278 bases at least Q30
                Consensus quality: 80916 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 44 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1519: contig of 1519 bp in length
            *     1520     1619: gap of unknown length
            *     1620     3041: contig of 1422 bp in length
            *     3042     3141: gap of unknown length
            *     3142     4192: contig of 1051 bp in length
            *     4193     4292: gap of unknown length
            *     4293     6074: contig of 1782 bp in length
            *     6075     6174: gap of unknown length
            *     6175     7346: contig of 1172 bp in length
            *     7347     7446: gap of unknown length
            *     7447     8693: contig of 1247 bp in length
            *     8694     8793: gap of unknown length
            *     8794    10240: contig of 1447 bp in length
            *    10241    10340: gap of unknown length
            *    10341    12235: contig of 1895 bp in length
            *    12236    12335: gap of unknown length
            *    12336    13390: contig of 1055 bp in length
            *    13391    13490: gap of unknown length
            *    13491    14571: contig of 1081 bp in length
            *    14572    14671: gap of unknown length
            *    14672    15864: contig of 1193 bp in length
            *    15865    15964: gap of unknown length
            *    15965    17316: contig of 1352 bp in length
            *    17317    17416: gap of unknown length
            *    17417    18656: contig of 1240 bp in length
            *    18657    18756: gap of unknown length
            *    18757    20056: contig of 1300 bp in length
            *    20057    20156: gap of unknown length
            *    20157    21931: contig of 1775 bp in length
            *    21932    22031: gap of unknown length
            *    22032    24000: contig of 1969 bp in length
            *    24001    24100: gap of unknown length
            *    24101    26222: contig of 2122 bp in length
            *    26223    26322: gap of unknown length
            *    26323    27382: contig of 1060 bp in length
            *    27383    27482: gap of unknown length
            *    27483    28666: contig of 1184 bp in length
            *    28667    28766: gap of unknown length
            *    28767    31151: contig of 2385 bp in length
            *    31152    31251: gap of unknown length
            *    31252    32815: contig of 1564 bp in length
            *    32816    32915: gap of unknown length
            *    32916    34953: contig of 2038 bp in length
            *    34954    35053: gap of unknown length
            *    35054    36696: contig of 1643 bp in length
            *    36697    36796: gap of unknown length
            *    36797    39079: contig of 2283 bp in length
            *    39080    39179: gap of unknown length
            *    39180    41478: contig of 2299 bp in length
            *    41479    41578: gap of unknown length
            *    41579    43648: contig of 2070 bp in length
            *    43649    43748: gap of unknown length
            *    43749    45639: contig of 1891 bp in length
            *    45640    45739: gap of unknown length
            *    45740    47441: contig of 1702 bp in length
            *    47442    47541: gap of unknown length
            *    47542    49632: contig of 2091 bp in length
            *    49633    49732: gap of unknown length
            *    49733    52604: contig of 2872 bp in length
            *    52605    52704: gap of unknown length
            *    52705    56288: contig of 3584 bp in length
            *    56289    56388: gap of unknown length
            *    56389    58558: contig of 2170 bp in length
            *    58559    58658: gap of unknown length
            *    58659    61781: contig of 3123 bp in length
            *    61782    61881: gap of unknown length
            *    61882    64670: contig of 2789 bp in length
            *    64671    64770: gap of unknown length
            *    64771    69117: contig of 4347 bp in length
            *    69118    69217: gap of unknown length
            *    69218    72604: contig of 3387 bp in length
            *    72605    72704: gap of unknown length
            *    72705    76341: contig of 3637 bp in length
            *    76342    76441: gap of unknown length
            *    76442    80834: contig of 4393 bp in length
            *    80835    80934: gap of unknown length
            *    80935    84092: contig of 3158 bp in length
```

```
            *    84093    84192: gap of unknown length
            *    84193    89965: contig of 5773 bp in length
            *    89966    90065: gap of unknown length
            *    90066    93596: contig of 3531 bp in length
            *    93597    93696: gap of unknown length
            *    93697    99934: contig of 6238 bp in length
            *    99935   100034: gap of unknown length
            *   100035   106684: contig of 6650 bp in length
            *   106685   106784: gap of unknown length
            *   106785   115667: contig of 8883 bp in length.
FEATURES             Location/Qualifiers
     source          1..115667
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-421M5"
BASE COUNT    29660 a  25606 c  26021 g  29956 t   4424 others
ORIGIN

Alignment Scores:
Pred. No.:              1.51e-07       Length:         115667
Score:                  18.00          Matches:        18
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            5.50%          Indels:         0
DB:                     2              Gaps:           0

US-09-357-675C-21 (1-327) x AC125857 (1-115667)

Qy     76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 93
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36950 AGACTGGGCGCTTGCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCACGA 36897

RESULT 15
AC105589
LOCUS       AC105589               181583 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
            ***, 67 unordered pieces.
ACCESSION   AC105589
VERSION     AC105589.2  GI:21736456
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 181583)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 181583)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18092812.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNGI
                Center clone name: CH230-242B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 112684 bases at least Q40
                Consensus quality: 120012 bases at least Q30
                Consensus quality: 127587 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 67 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1312: contig of 1312 bp in length
            *     1313     1412: gap of unknown length
            *     1413     2762: contig of 1350 bp in length
            *     2763     2862: gap of unknown length
            *     2863     4412: contig of 1550 bp in length
            *     4413     4512: gap of unknown length
            *     4513     5532: contig of 1020 bp in length
            *     5533     5632: gap of unknown length
            *     5633     6814: contig of 1182 bp in length
            *     6815     6914: gap of unknown length
            *     6915     8099: contig of 1185 bp in length
            *     8100     8199: gap of unknown length
            *     8200     9786: contig of 1587 bp in length
            *     9787     9886: gap of unknown length
            *     9887    11574: contig of 1688 bp in length
            *    11575    11674: gap of unknown length
            *    11675    12816: contig of 1142 bp in length
            *    12817    12916: gap of unknown length
            *    12917    14067: contig of 1151 bp in length
```

```
*    14068    14167: gap of unknown length
*    14168    15370: contig of 1203 bp in length
*    15371    15470: gap of unknown length
*    15471    16926: contig of 1456 bp in length
*    16927    17026: gap of unknown length
*    17027    18088: contig of 1062 bp in length
*    18089    18188: gap of unknown length
*    18189    19315: contig of 1127 bp in length
*    19316    19415: gap of unknown length
*    19416    20772: contig of 1357 bp in length
*    20773    20872: gap of unknown length
*    20873    21937: contig of 1065 bp in length
*    21938    22037: gap of unknown length
*    22038    23147: contig of 1110 bp in length
*    23148    23247: gap of unknown length
*    23248    24803: contig of 1556 bp in length
*    24804    24903: gap of unknown length
*    24904    26111: contig of 1208 bp in length
*    26112    26211: gap of unknown length
*    26212    27647: contig of 1436 bp in length
*    27648    27747: gap of unknown length
*    27748    30257: contig of 2510 bp in length
*    30258    30357: gap of unknown length
*    30358    31725: contig of 1368 bp in length
*    31726    31825: gap of unknown length
*    31826    32912: contig of 1087 bp in length
*    32913    33012: gap of unknown length
*    33013    34593: contig of 1581 bp in length
*    34594    34693: gap of unknown length
*    34694    36338: contig of 1645 bp in length
*    36339    36438: gap of unknown length
*    36439    38139: contig of 1701 bp in length
*    38140    38239: gap of unknown length
*    38240    39349: contig of 1110 bp in length
*    39350    39449: gap of unknown length
*    39450    40968: contig of 1519 bp in length
*    40969    41068: gap of unknown length
*    41069    42068: contig of 1000 bp in length
*    42069    42168: gap of unknown length
*    42169    43869: contig of 1701 bp in length
*    43870    43969: gap of unknown length
*    43970    46141: contig of 2172 bp in length
*    46142    46241: gap of unknown length
*    46242    48158: contig of 1917 bp in length
*    48159    48258: gap of unknown length
*    48259    50613: contig of 2355 bp in length
*    50614    50713: gap of unknown length
*    50714    52725: contig of 2012 bp in length
*    52726    52825: gap of unknown length
*    52826    54534: contig of 1709 bp in length
*    54535    54634: gap of unknown length
*    54635    56455: contig of 1821 bp in length
*    56456    56555: gap of unknown length
*    56556    58494: contig of 1939 bp in length
*    58495    58594: gap of unknown length
*    58595    60496: contig of 1902 bp in length
*    60497    60596: gap of unknown length
*    60597    62628: contig of 2032 bp in length
*    62629    62728: gap of unknown length
*    62729    65060: contig of 2332 bp in length
*    65061    65160: gap of unknown length
*    65161    66543: contig of 1383 bp in length
*    66544    66643: gap of unknown length
*    66644    68412: contig of 1769 bp in length
*    68413    68512: gap of unknown length
*    68513    71585: contig of 3073 bp in length
*    71586    71685: gap of unknown length
*    71686    74575: contig of 2890 bp in length
*    74576    74675: gap of unknown length
*    74676    77865: contig of 3190 bp in length
*    77866    77965: gap of unknown length
*    77966    80442: contig of 2477 bp in length
*    80443    80542: gap of unknown length
*    80543    83108: contig of 2566 bp in length
*    83109    83208: gap of unknown length
*    83209    86895: contig of 3687 bp in length
*    86896    86995: gap of unknown length
*    86996    91038: contig of 4043 bp in length
*    91039    91138: gap of unknown length
*    91139    94412: contig of 3274 bp in length
*    94413    94512: gap of unknown length
*    94513    96959: contig of 2447 bp in length
*    96960    97059: gap of unknown length
*    97060    98988: contig of 1929 bp in length
*    98989    99088: gap of unknown length
*    99089   101633: contig of 2545 bp in length
*   101634   101733: gap of unknown length
*   101734   105899: contig of 4166 bp in length

Alignment Scores:
Pred. No.:                2.27e-07        Length:          181583
Score:                    18.00           Matches:         18
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              5.50%           Indels:          0
DB:                       2               Gaps:            0

US-09-357-675C-21 (1-327) x AC105589 (1-181583)

Qy      76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 93
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133573 AGACTGGGCGCTTGCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCACGA 133626


Search completed: December  8, 2002, 11:33:16
Job time : 2957 secs
```

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2002, 10:30:41 ; Search time 255 Seconds
(without alignments)
2887.857 Million cell updates/sec

Title: US-09-357-675C-21
Perfect score: 327
Sequence: 1 MLGFITRPPHRFLSLLCPGL.....LPVFQHRRPDLYGNLGHPLS 327

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09357675/runat_03122002_142235_20387/app_query.fasta_1.519
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09357675_@CGN_1_1_125_@runat_03122002_142235_20387 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		326	99.7	1416	21	AAZ46101	Nucleotide sequenc
2		253	77.4	1382	21	AAF16257	Human prostate can
3		153	46.8	1203	20	AAX30398	DNA encoding a hum
c 4		63	19.3	377	24	ABK39058	cDNA encoding lung
5		23	7.0	1214	24	AAD25458	Xenopus laevis Nit
6		18	5.5	539	22	AAK61236	Human immune/haema
7		13	4.0	604	24	ABQ66352	Arabidopsis thalia
8		11	3.4	1495	23	ABL12225	Drosophila melanog
9		11	3.4	3548	23	ABL12224	Drosophila melanog
10		11	3.4	5692	23	ABL20222	Drosophila melanog
11		9	2.8	242	24	ABN23783	Human ORFX polynuc
c 12		9	2.8	6597	22	AAH19580	Human polynucleoti
c 13		9	2.8	6597	24	ABK84618	Human cDNA differe
c 14		9	2.8	69936	21	AAA81479	N. meningitidis pa
15		9	2.8	349980	21	AAF21607	Neisseria meningit
16		9	2.8	1437668	21	AAA81490	N. meningitidis B
17		8	2.4	44	24	AAS18766	PCR primer #7 used
18		8	2.4	56	21	AAC65684	Human BPI PCR prim
c 19		8	2.4	90	21	AAC15536	Human secreted pro
c 20		8	2.4	115	22	AAF76379	Escherichia coli J
c 21		8	2.4	115	24	AAD27662	Synthetic DNA targ
c 22		8	2.4	250	23	ABK42816	Genomic sequence #
c 23		8	2.4	280	22	ABA19047	Human nervous syst
c 24		8	2.4	291	21	AAC15522	Human secreted pro
c 25		8	2.4	314	24	ABN21264	Human ORFX polynuc
c 26		8	2.4	330	23	ABL25191	Drosophila melanog
c 27		8	2.4	351	22	AAF76378	Escherichia coli J
c 28		8	2.4	351	24	AAD27663	Single-stranded DN
c 29		8	2.4	380	24	ABK63140	Rat sequence diffe
c 30		8	2.4	403	22	AAK78684	Human immune/haema
c 31		8	2.4	441	24	ABK46564	DNA encoding human
c 32		8	2.4	470	22	AAI92703	Human polynucleoti
c 33		8	2.4	475	20	AAZ25127	Wheat tryptophan s
34		8	2.4	482	22	AAK70899	Human immune/haema
c 35		8	2.4	494	23	ABV09854	Human prostate exp
36		8	2.4	509	22	AAK76654	Human immune/haema
37		8	2.4	574	22	AAK64697	Human immune/haema
38		8	2.4	601	22	AAF94030	Primer specific fo
39		8	2.4	639	23	ABL23305	Drosophila melanog
40		8	2.4	683	24	ABQ38874	Oligonucleotide fo
c 41		8	2.4	683	24	ABQ38875	Oligonucleotide fo
c 42		8	2.4	708	22	AAH04672	Human cDNA clone (
43		8	2.4	739	22	AAF94008	Primer specific fo
44		8	2.4	787	22	AAI96497	Human neuroblastom
45		8	2.4	859	20	AAX97989	Human secreted pro

ALIGNMENTS

RESULT 1
AAZ46101
ID AAZ46101 standard; DNA; 1416 BP.
XX
AC AAZ46101;
XX
DT 05-MAY-2000 (first entry)
XX
DE Nucleotide sequence of the coding region of NIT1 gene.
XX
KW NIT1 gene; nitrilase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW FRA3B; cancer; genome allele inactivation; ss.
XX
OS Homo sapiens.
OS Mus sp.
OS Drosophila melanogaster.
OS Caenorhabditis elegans.
XX

```
FH   Key             Location/Qualifiers
FT   CDS             3..1415
FT                   /*tag=  a
FT                   /product= (pos: 18..20, aa: Xaa)
FT                   /product= (pos: 1179..1181, aa: Gln)
FT                   /product= (pos: 1182..1184, aa: Ala)
FT                   /note= "contains 7 internal stop codons; Xaa is an
FT                          unspecified amino acid"
XX
PN   WO200003685-A2.
XX
PD   27-JAN-2000.
XX
PF   20-JUL-1999;   99WO-US16366.
XX
PR   20-JUL-1998;   98US-0093350.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce CM;
XX
DR   WPI; 2000-171195/15.
DR   P-PSDB; AAY68739.
XX
PT   Novel nitrilase homologs used as diagnostic and therapeutic reagents
PT   for the detection and treatment of cancer  -
XX
PS   Claim 6; Fig 6; 25pp; English.
XX
CC   The present sequence represents the coding region of human, murine,
CC   Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC   human and mouse NIT1 genes are members of an uncharacterised
CC   mammalian gene family with homology to bacterial and plant nitrilases.
CC   The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC   for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC   In mouse and humans, FHIT and NIT are encoded by two different genes,
CC   localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC   human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC   chromosomal fragile site FRA3B, is often altered in most common forms
CC   of human cancer. The Nit1 protein overcomes the mutated inactivation
CC   of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC   and analogues of them, and antibodies are used as diagnostic and
CC   therapeutic reagents for the detection and treatment of cancers.
XX
SQ   Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

Alignment Scores:
Pred. No.:              0           Length:         1416
Score:                  326.00      Matches:        326
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
Query Match:            99.69%      Indels:         0
DB:                     21          Gaps:           0

US-09-357-675C-21 (1-327) x AAZ46101 (1-1416)

Qy      2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeuArg 21
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGG 170

Qy     22 IleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSer 41
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 ATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCT 230

Qy     42 SerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 TCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAA 290

Qy     62 GlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeu 81
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 CAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTG 350

Qy     82 AlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTG 410

Qy    102 SerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGly 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 TCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGA 470

Qy    122 LeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLys 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 CTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAGAAA 530

Qy    142 IleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArgLys 161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 ATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGGAAG 590

Qy    162 ThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThr 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 ACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATGTGTGAAAGCAACTCTACC 650

Qy    182 MetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 ATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTAGCT 710

Qy    202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGlu 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 GTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAG 770

Qy    222 IleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeu 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 ATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTGTTG 830

Qy    242 LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArg 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 CTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGACGC 890

Qy    262 HisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVal 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891 CACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTG 950

Qy    282 ValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 GTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTATCTG 1010

Qy    302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsn 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 CGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGCAAT 1070

Qy    322 LeuGlyHisProLeuSer 327
          |||||||||||||||||
Db   1071 CTGGGTCACCCACTGTCT 1088

RESULT 2
AAF16257
ID   AAF16257 standard; cDNA; 1382 BP.
XX
AC   AAF16257;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
```

```
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB57054.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 1124; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

Alignment Scores:
Pred. No.:              1e-247          Length:         1382
Score:                  253.00          Matches:        326
Percent Similarity:     99.69%          Conservative:   0
Best Local Similarity:  99.69%          Mismatches:     1
Query Match:            77.37%          Indels:         1
DB:                     21              Gaps:           0

US-09-357-675C-21 (1-327) x AAF16257 (1-1382)

Qy       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeu 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTC 136

Qy      21 ArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 CGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCC 196

Qy      41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLys 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 TCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAG 256

Qy      61 GlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CAACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGC 316

Qy      81 LeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

Qy     101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCys 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGT 436

Qy     121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 GGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAG 496

Qy     141 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 AAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGG 556

Qy     161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
           ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db     557 AAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCT-ATGTGTGAAAGCAACTCT 615

Qy     181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     616 ACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 675

Qy     201 AlaValCysTyrAspMetArgProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     676 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCA 735

Qy     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     736 GAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTG 795

Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGly 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     796 TTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGA 855

Qy     261 ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     856 CGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACA 915

Qy     281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     916 GTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 975

Qy     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     976 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGC 1035

Qy     321 AsnLeuGlyHisProLeuSer 327
           |||||||||||||||||||||
Db    1036 AATCTGGGTCACCCACTGTCT 1056

RESULT 3
AAX30398
ID   AAX30398 standard; DNA; 1203 BP.
XX
AC   AAX30398;
XX
DT   14-MAY-1999  (first entry)
XX
DE   DNA encoding a human secreted protein.
XX
KW   Secreted protein; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder;
KW   CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW   renal disease; diabetes; inflammation; allergy; ischemic shock;
KW   Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW   prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9907891-A1.
XX
PD   18-FEB-1999.
XX
PF   04-AUG-1998;   98WO-US16235.
XX
PR   19-AUG-1997;   97US-0056732.
PR   05-AUG-1997;   97US-0054798.
PR   05-AUG-1997;   97US-0054803.
PR   05-AUG-1997;   97US-0054804.
PR   05-AUG-1997;   97US-0054806.
PR   05-AUG-1997;   97US-0054807.
PR   05-AUG-1997;   97US-0054808.
```

```
PR   05-AUG-1997;   97US-0054809.
PR   05-AUG-1997;   97US-0055309.
PR   05-AUG-1997;   97US-0055310.
PR   05-AUG-1997;   97US-0055312.
PR   05-AUG-1997;   97US-0055386.
PR   05-AUG-1997;   97US-0055311.
PR   18-AUG-1997;   97US-0055970.
PR   18-AUG-1997;   97US-0055986.
PR   19-AUG-1997;   97US-0056365.
PR   19-AUG-1997;   97US-0056366.
PR   19-AUG-1997;   97US-0056557.
PR   19-AUG-1997;   97US-0056370.
PR   19-AUG-1997;   97US-0056371.
PR   19-AUG-1997;   97US-0056563.
PR   19-AUG-1997;   97US-0056731.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Ebner R,  Ferrie AM,  Greene JM,  Janat F,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR,  Young PE,  Yu G;
XX
DR   WPI; 1999-167452/14.
DR   P-PSDB; AAY10877.
XX
PT   New isolated human genes encoding secreted polypeptides - useful for
PT   diagnosis and treatment of pathalogical diseases
XX
PS   Claim 3; Page 265-266; 331pp; English.
XX
CC   The specification describes secreted proteins and their corresponding
CC   polynucleotides which are useful for preventing, treating or ameliorating
CC   medical conditions, e.g. by protein or gene therapy. Pathological
CC   conditions can also be diagnosed by determining the amount of the
CC   secreted polypeptides in a sample or by determining the presence of
CC   mutations in the polynucleotides. Specific uses are described for each
CC   of the products, based on which tissues they are most highly
CC   expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC   and cognitive disorders, schizophrenia, cardiovascular disorders,
CC   prostate diseases, asthma, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC   thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
XX
SQ   Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

Alignment Scores:
Pred. No.:              6.74e-146      Length:          1203
Score:                  153.00         Matches:         205
Percent Similarity:     99.03%         Conservative:    0
Best Local Similarity:  99.03%         Mismatches:      2
Query Match:            46.79%         Indels:          2
DB:                     20             Gaps:            0

US-09-357-675C-21 (1-327) x AAX30398 (1-1203)

Qy    121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 GGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAG 62

Qy    141 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     63 AAAATCTACAATTGTCACGTGCTGCTGAACAG-AAAGGGGCAGTAGTGGCCACTTACAGG 121

Qy    161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
          |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db    122 AAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCT-ATGTGTGAAAGCAACTCT 180

Qy    181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    181 ACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 240

Qy    201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCA 300

Qy    221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTG 360

Qy    241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGly 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGA 420

Qy    261 ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACA 480

Qy    281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 540

Qy    301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGC 600

Qy    321 AsnLeuGlyHisProLeuSer 327
          |||||||||||||||||||||
Db    601 AATCTGGGTCACCCACTGTCT 621

RESULT 4
ABK39058/c
ID   ABK39058 standard; cDNA; 377 BP.
XX
AC   ABK39058;
XX
DT   21-MAY-2002  (first entry)
XX
DE   cDNA encoding lung tumour protein clone R0130:B11.
XX
KW   Lung tumour; cancer; T cell; immune response stimulator;
KW   cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200204514-A2.
XX
PD   17-JAN-2002.
XX
PF   10-JUL-2001; 2001WO-US22058.
XX
PR   11-JUL-2000; 2000US-0614124.
PR   29-AUG-2000; 2000US-0651563.
PR   08-SEP-2000; 2000US-0658824.
PR   26-SEP-2000; 2000US-0671325.
PR   06-OCT-2000; 2000US-0677419.
PR   30-OCT-2000; 2000US-0702705.
PR   13-DEC-2000; 2000US-0736457.
PR   03-MAY-2001; 2001US-0849626.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Watanabe Y,  Henderson RA,  Johnson JC,  Retter MW;
PI   Marnerakis M,  Carter D,  Fanger GR,  Vedvick TS,  Bangur CS;
PI   Mcnabb A,  Wang A,  Fanger N,  Switzer A,  Mcneill PD,  Clapper JD;
XX
DR   WPI; 2002-164634/21.
XX
PT   Novel polynucleotide encoding a lung tumour polypeptide useful for
PT   stimulating and/or expanding T cells specific for a tumour protein  -
XX
PS   Claim 1; SEQ ID No 1096; 223pp; English.
XX
```

```
CC   The invention describes an isolated polynucleotide and polypeptide
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein for determining the presence of a cancer in a patient. A
CC   composition containing the polynucleotide and/or polypeptide is useful
CC   for treating a lung cancer in a patient. The polypeptide is useful for
CC   removing tumour cells from a biological sample. The polynucleotide is
CC   also useful as probe or primer to detect the level of mRNA encoding a
CC   tumour protein. This sequence encodes a lung tumour associated protein
CC   or protein fragment, described in the method of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 377 BP; 94 A; 100 C; 112 G; 71 T; 0 other;

Alignment Scores:
Pred. No.:              1.08e-54       Length:          377
Score:                  63.00          Matches:         124
Percent Similarity:     98.41%         Conservative:    0
Best Local Similarity:  98.41%         Mismatches:      1
Query Match:            19.27%         Indels:          2
DB:                     24             Gaps:            0

US-09-357-675C-21 (1-327) x ABK39058 (1-377)

Qy       8 ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 27
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy      28 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuPro 47
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGGCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy      48 LeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy      68 AlaGlu-LeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAl 87
           |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGG-GCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGC 139

Qy      87 aPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGl 107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGG 79

Qy     107 yLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGG 19

Qy     127 yGlyPheHisGluArg 132
           |||||||||||||||
Db      18 TGGTTTCCATGAGCGT 3

RESULT 5
AAD25458
ID   AAD25458 standard; cDNA; 1214 BP.
XX
AC   AAD25458;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Xenopus laevis Nit1 cDNA.
XX
KW   Frog; Nit1; cytostatic; neuroprotective; cellular pathway; therapy;
KW   apoptosis; proliferative disorder; degenerative disease; ss.
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
FT   CDS             180..1046
FT                   /*tag= a
FT                   /product= "Frog Nit1 protein"
XX
PN   WO200187958-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US15664.
XX
PR   16-MAY-2000; 2000US-204713P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Croce C,  Brenner C,  Pekarski Y;
XX
DR   WPI; 2002-082984/11.
DR   P-PSDB; AAE15792.
XX
PT   Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
PT   useful to find molecules that mimic or antagonize Fhit interaction for
PT   the treatment of proliferative or degenerative diseases -
XX
PS   Claim 21; Page 60; 61pp; English.
XX
CC   The invention relates to isolated nucleic acids comprising a fully
CC   defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC   mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion
CC   proteins in invertebrates and as separate polypeptides in vertebrates.
CC   Nit and Fhit interact physically and functions in same cellular pathways.
CC   Molecules which bind Nit2 and mimic or antagonise Fhit interaction are
CC   used to treat diseases in which activity of Nit2 protein is altered in
CC   a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC   treat proliferative disorders, whilst Fhit antagonists promote cell
CC   proliferation and are particularly useful to treat degenerative disease.
CC   The present sequence is frog Nit1 cDNA.
XX
SQ   Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

Alignment Scores:
Pred. No.:              1.93e-13       Length:          1214
Score:                  23.00          Matches:         23
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            7.03%          Indels:          0
DB:                     24             Gaps:            0

US-09-357-675C-21 (1-327) x AAD25458 (1-1214)

Qy     236 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAla 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     765 GCACATTGGGAGGTGTTGCTGAGAGCCCGTGCCATAGAAACCCAGTGCTACGTAGTTGCA 824

Qy     256 AlaAlaGln 258
           ||||||||
Db     825 GCGGCACAG 833

RESULT 6
AAK61236
ID   AAK61236 standard; cDNA; 539 BP.
XX
AC   AAK61236;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6296.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
```

PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR P-PSDB; AAM88455.
XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 1; SEQ ID NO 6296; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

Alignment Scores:

Pred. No.:	1.08e-08	Length:	539
Score:	18.00	Matches:	18
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.50%	Indels:	0
DB:	22	Gaps:	0

US-09-357-675C-21 (1-327) x AAK61236 (1-539)

```
Qy       2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGly 19
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 CTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCKGGN 426
```

RESULT 7
ABQ66352
ID ABQ66352 standard; DNA; 604 BP.
XX
AC ABQ66352;
XX
DT 21-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana polynucleotide SEQ ID NO 929.
XX
KW Arabidopsis thaliana; thale cress; plant; transgenic; GMO; disease;
KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
KW insecticide; antibiotic; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002059663-A1.
XX
PD 16-MAY-2002.
XX
PF 26-JAN-2001; 2001US-0770149.
XX
PR 27-JAN-2000; 2000US-178506P.
XX
PA (GORL/) GORLACH J.
PA (ANYY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUYY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
DR WPI; 2002-479224/51.
XX
PT New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
PT useful e.g. for preparing transgenic plants with increased resistance
PT or altered metabolism -
XX
PS Claim 1; SEQ ID NO 929; 40pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acids (I) that hybridise under stringent
CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their
CC fragments. (I) are used to express the corresponding polypeptides (II) or
CC to produce genetically modified plant cells or transgenic plants, which
CC may have improved resistance to disease or stress, or altered
CC metabolic/biosynthetic pathways (for production of commercial,
CC nutritional or medicinal products), or generally any trait of interest,
CC or can be used to screen for biologically active agents (e.g. fungicides,
CC insecticides and antibiotics).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=999909770149.
XX
SQ Sequence 604 BP; 174 A; 113 C; 139 G; 176 T; 2 other;

Alignment Scores:

Pred. No.:	0.0015	Length:	604
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.98%	Indels:	0
DB:	24	Gaps:	0

US-09-357-675C-21 (1-327) x ABQ66352 (1-604)

```
Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 253
           |||||||||||||||||||||||||||||||||||||
Db     198 CTTCTTCGAGCCCGAGCAATTGAAACTCAATGTTATGTC 236
```

RESULT 8
ABL12225
ID ABL12225 standard; cDNA; 1495 BP.
XX
AC ABL12225;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB68122.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 31157; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

Alignment Scores:

Pred. No.:	0.398	Length:	1495
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.36%	Indels:	0
DB:	23	Gaps:	0

US-09-357-675C-21 (1-327) x ABL12225 (1-1495)

```
Qy    241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251
          ||||||||||||||||||||||||||||||||
Db    704 CTCCTGCGGGCCAGAGCCATAGAGACTCAATGC 736
```

RESULT 9
ABL12224
ID ABL12224 standard; cDNA; 3548 BP.
XX
AC ABL12224;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB68121.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 31154; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

Alignment Scores:

Pred. No.:	0.93	Length:	3548
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.36%	Indels:	0
DB:	23	Gaps:	0

US-09-357-675C-21 (1-327) x ABL12224 (1-3548)

```
Qy    241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251
          ||||||||||||||||||||||||||||||||
Db   1757 CTCCTGCGGGCCAGAGCCATAGAGACTCAATGC 1789
```

RESULT 10
ABL20222
ID ABL20222 standard; DNA; 5692 BP.
XX
AC ABL20222;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.
XX

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

Alignment Scores:
Pred. No.:                1.48          Length:         5692
Score:                    11.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              3.36%         Indels:         0
DB:                       23            Gaps:           0

US-09-357-675C-21 (1-327) x ABL20222 (1-5692)

Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251
           ||||||||||||||||||||||||||||||||
Db    3416 CTCCTGCGGGCCAGAGCCATAGAGACTCAATGC 3448

RESULT 11
ABN23783
ID   ABN23783 standard; cDNA; 242 BP.
XX
AC   ABN23783;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:16043.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP08031.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 16043; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 242 BP; 78 A; 48 C; 52 G; 63 T; 1 other;

Alignment Scores:
Pred. No.:                7.26          Length:         242
Score:                    9.00          Matches:        9
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              2.75%         Indels:         0
DB:                       24            Gaps:           0

US-09-357-675C-21 (1-327) x ABN23783 (1-242)

Qy     203 CysTyrAspMetArgPheProGluLeu 211
           |||||||||||||||||||||||||||
Db     199 TGCTATGACATGAGGTTTCCGGAACTC 225

RESULT 12
AAH19580/c
ID   AAH19580 standard; cDNA; 6597 BP.
XX
AC   AAH19580;
XX
DT   26-JUL-2001  (first entry)
XX
DE   Human polynucleotide #3 expressed in intraabdominal adipose tissue.
XX
KW   Human; intraabdominal adipose tissue; gene expression; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..6367
FT                   /*tag=  a
XX
PN   JP2001008699-A.
XX
PD   16-JAN-2001.
XX
PF   30-JUN-1999;   99JP-0185737.
XX
PR   30-JUN-1999;   99JP-0185737.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 2001-275911/29.
DR   P-PSDB; AAB97070.
XX
PT   Assuming the intraabdominal adipose tissue amount, comprising relating
PT   the expression level of a gene expression product to the amount -
```

XX
PS Claim 1; Page 15-23; 33pp; Japanese.
XX
CC The present sequence is provided in a specification relating to a
CC method for predicting the amount of intraabdominal adipose tissue.
CC The method involves relating the amount of adipose tissue to the
CC expression level of at least one gene transcription product which
CC has a 491 or 2090 residue amino acid sequence and is encoded by a
CC 2376 or 2090 base pair sequence, or which can hybridise to a 2385
CC or the 2090 base pair sequence under stringent conditions. The
CC method includes a step of deriving the area value of the
CC intraabdominal adipose tissue at the cross section of abdominal
CC navel, from the expression level of a gene transcription product
CC in a sample of the intraabdominal adipose tissue. The method can
CC be used for predicting the amount of intraabdominal adipose
CC tissue irrespective of the amount of subcutaneous adipose tissue.
CC The present sequence encodes a gene transcription product whose
CC expression level may be measured as part of this method.
XX
SQ Sequence 6597 BP; 1590 A; 1870 C; 1549 G; 1588 T; 0 other;

Alignment Scores:

Pred. No.:	186	Length:	6597
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.75%	Indels:	0
DB:	22	Gaps:	0

US-09-357-675C-21 (1-327) x AAH19580 (1-6597)

```
Qy    212 SerLeuAlaLeuAlaGlnAlaGlyAla 220
          |||||||||||||||||||||||||||
Db   2660 TCTCTCGCTCTGGCACAAGCAGGTGCC 2634
```

RESULT 13
ABK84618/c
ID ABK84618 standard; cDNA; 6597 BP.
XX
AC ABK84618;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1189.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 1189; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6597 BP; 1590 A; 1870 C; 1549 G; 1588 T; 0 other;

Alignment Scores:

Pred. No.:	186	Length:	6597
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.75%	Indels:	0
DB:	24	Gaps:	0

US-09-357-675C-21 (1-327) x ABK84618 (1-6597)

```
Qy    212 SerLeuAlaLeuAlaGlnAlaGlyAla 220
          |||||||||||||||||||||||||||
Db   2660 TCTCTCGCTCTGGCACAAGCAGGTGCC 2634
```

RESULT 14
AAA81479/c
ID AAA81479 standard; DNA; 69936 BP.
XX
AC AAA81479;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX

PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 547-567; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

Alignment Scores:

Pred. No.:	1.89e+03	Length:	69936
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.75%	Indels:	0
DB:	21	Gaps:	0

US-09-357-675C-21 (1-327) x AAA81479 (1-69936)

```
Qy    250 GlnCysTyrValValAlaAlaAlaGln 258
          |||||||||||||||||||||||||||
Db  26543 CAATGTTACGTCGTGGCGGCGGCACAG 26517
```

RESULT 15
AAF21607
ID AAF21607 standard; DNA; 349980 BP.
XX
AC AAF21607;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Alignment Scores:

Pred. No.:	9.2e+03	Length:	349980
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.75%	Indels:	0
DB:	21	Gaps:	0

US-09-357-675C-21 (1-327) x AAF21607 (1-349980)

```
Qy    250 GlnCysTyrValValAlaAlaAlaGln 258
          |||||||||||||||||||||||||||
Db 154644 CAATGTTACGTCGTGGCGGCGGCACAG 154670
```

Search completed: December 8, 2002, 10:45:31

Job time : 368 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2002, 10:38:31 ; Search time 52 Seconds
(without alignments)
1928.525 Million cell updates/sec

Title: US-09-357-675C-21
Perfect score: 327
Sequence: 1 MLGFITRPPHRFLSLLCPGL.....LPVFQHRRPDLYGNLGHPLS 327

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09357675/runat_03122002_142236_20423/app_query.fasta_1.519
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09357675_@CGN_1_1_17_@runat_03122002_142236_20423 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		8	2.4	56	1	US-08-173-968-6	Sequence 6, Appli
2		8	2.4	56	3	US-07-801-814D-6	Sequence 6, Appli
3	c	8	2.4	115	4	US-09-660-877-18	Sequence 18, Appl
4	c	8	2.4	351	4	US-09-660-877-17	Sequence 17, Appl
5	c	8	2.4	864	4	US-09-439-261-12	Sequence 12, Appl
6	c	8	2.4	864	4	US-09-227-613-13	Sequence 13, Appl
7	c	8	2.4	918	4	US-09-439-261-5	Sequence 5, Appli
8	c	8	2.4	918	4	US-09-227-613-5	Sequence 5, Appli
9	c	8	2.4	960	4	US-09-439-261-36	Sequence 36, Appl
10	c	8	2.4	960	4	US-09-227-613-35	Sequence 35, Appl
11	c	8	2.4	990	4	US-09-439-261-35	Sequence 35, Appl
12	c	8	2.4	990	4	US-09-227-613-34	Sequence 34, Appl
13	c	8	2.4	1335	4	US-09-439-261-1	Sequence 1, Appli
14	c	8	2.4	1335	4	US-09-227-613-1	Sequence 1, Appli
15	c	8	2.4	1508	4	US-09-039-046-1	Sequence 1, Appli
16	c	8	2.4	1881	4	US-09-553-498-1	Sequence 1, Appli
17	c	8	2.4	1881	4	US-09-553-498-3	Sequence 3, Appli
18	c	8	2.4	1881	4	US-09-618-869-1	Sequence 1, Appli
19	c	8	2.4	1881	4	US-09-618-869-3	Sequence 3, Appli
20	c	8	2.4	1946	2	US-08-755-584-1	Sequence 1, Appli
21	c	8	2.4	1946	3	US-09-192-611-1	Sequence 1, Appli
22	c	8	2.4	2159	4	US-09-039-046-4	Sequence 4, Appli
23	c	8	2.4	3300	1	US-08-194-290-6	Sequence 6, Appli
24	c	8	2.4	3300	2	US-08-614-377A-6	Sequence 6, Appli
25	c	8	2.4	3300	4	US-09-142-648B-6	Sequence 6, Appli
26		8	2.4	3350	1	US-08-247-946A-2	Sequence 2, Appli
27		8	2.4	3350	5	PCT-US95-06420-2	Sequence 2, Appli
28		8	2.4	4085	1	US-08-072-574-7	Sequence 7, Appli
29		8	2.4	4085	1	US-08-486-270-7	Sequence 7, Appli
30		8	2.4	4085	3	US-08-367-264-7	Sequence 7, Appli
31		8	2.4	4085	4	US-09-153-757-7	Sequence 7, Appli
32		8	2.4	4181	1	US-08-072-574-9	Sequence 9, Appli
33		8	2.4	4181	1	US-08-486-270-9	Sequence 9, Appli
34		8	2.4	4181	3	US-08-367-264-9	Sequence 9, Appli
35		8	2.4	4181	4	US-09-153-757-9	Sequence 9, Appli
36		8	2.4	4207	4	US-08-660-148-1	Sequence 1, Appli
37		8	2.4	4207	4	US-08-660-148-3	Sequence 3, Appli
38		8	2.4	4303	4	US-08-660-148-4	Sequence 4, Appli
39		8	2.4	4303	4	US-08-660-148-6	Sequence 6, Appli
40		8	2.4	4326	2	US-08-852-807-12	Sequence 12, Appl
41	c	8	2.4	4817	1	US-07-951-715A-18	Sequence 18, Appl
42	c	8	2.4	4817	2	US-08-459-448A-18	Sequence 18, Appl
43	c	8	2.4	4817	3	US-08-459-595A-18	Sequence 18, Appl
44	c	8	2.4	4817	3	US-08-459-504B-18	Sequence 18, Appl
45	c	8	2.4	4817	3	US-08-459-444-18	GENERAL INFORMA

ALIGNMENTS

RESULT 1
US-08-173-968-6
; Sequence 6, Application US/08173968
; Patent No. 5576292
; GENERAL INFORMATION:
; APPLICANT: Elsbach, Peter
; APPLICANT: Weiss, Jerrold
; TITLE OF INVENTION: Biologically Active
; TITLE OF INVENTION: Bactericidal/Permeability-increasing Protein Fragments
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howard M. Frankfort
; STREET: 805 Third Ave.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,968
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/754,204
; FILING DATE: 26-AUG-1991
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Frankfort, Howard M.
; REGISTRATION NUMBER: 32,613
; REFERENCE/DOCKET NUMBER: 5969/05982US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700

```
;      TELEFAX:  (212) 753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  56 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (synthetic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..56
;      OTHER INFORMATION:  note=This sequence can be found
;      OTHER INFORMATION:  in the application, as filed on page 25, line 28.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Ooi, C. E. et al.
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  262
;      PAGES:  14891-14894
;      DATE:  1987
US-08-173-968-6

Alignment Scores:
Pred. No.:              5.49          Length:         56
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     1             Gaps:           0

US-09-357-675C-21 (1-327) x US-08-173-968-6 (1-56)

Qy    286 SerGluGlyProGlyLeuCysLeu 293
          |||||||||||||||||||||||
Db     30 TCAGAAGGGCCTGGATTATGCCTC 53

RESULT 2
US-07-801-814D-6
; Sequence 6, Application US/07801814D
; Patent No. 6132775
;  GENERAL INFORMATION:
;    APPLICANT:  Elsbach, Peter
;    APPLICANT:  Weiss, Jerrold
;    TITLE OF INVENTION:  BIOLOGICALLY ACTIVE
;    TITLE OF INVENTION:  BACTERICIDAL/PERMEABILITY-INCREASING
;    TITLE OF INVENTION:  PROTEIN FRAGMENTS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby P.C.
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch,
;      MEDIUM TYPE:  360 Kb storage
;      COMPUTER:  IBM or IBM-compatible
;      OPERATING SYSTEM:  PC/MS-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/801,814D
;      FILING DATE:  19911204
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Emily Miao
;      REGISTRATION NUMBER:  35,285
;      REFERENCE/DOCKET NUMBER:  5986/13611-US3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  56
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  This sequence can be
;      OTHER INFORMATION:  found in the application, as filed, on
;      OTHER INFORMATION:  page 23, line 6.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Ooi, C. E. et al.
;      TITLE:
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  262
;      ISSUE:
;      PAGES:  14891-14894
;      DATE:  1987
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:  1-56
US-07-801-814D-6

Alignment Scores:
Pred. No.:              5.49          Length:         56
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     3             Gaps:           0

US-09-357-675C-21 (1-327) x US-07-801-814D-6 (1-56)

Qy    286 SerGluGlyProGlyLeuCysLeu 293
          |||||||||||||||||||||||
Db     30 TCAGAAGGGCCTGGATTATGCCTC 53

RESULT 3
US-09-660-877-18/c
; Sequence 18, Application US/09660877
; Patent No. 6251639
; GENERAL INFORMATION:
;  APPLICANT:  Kurn, N.
;  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR LINEAR
;  TITLE OF INVENTION:  ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES
;  FILE REFERENCE:  492692000100
;  CURRENT APPLICATION NUMBER: US/09/660,877
;  CURRENT FILING DATE:  2000-09-13
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 115
;   TYPE: DNA
;   ORGANISM: Synthetic primer
;   FEATURE:
;   OTHER INFORMATION: IA013
US-09-660-877-18

Alignment Scores:
Pred. No.:              10.8          Length:         115
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
```

```
Query Match:            2.45%          Indels:        0
DB:                     4              Gaps:          0

US-09-357-675C-21 (1-327) x US-09-660-877-18 (1-115)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 4
US-09-660-877-17/c
; Sequence 17, Application US/09660877
; Patent No. 6251639
; GENERAL INFORMATION:
;  APPLICANT:  Kurn, N.
;  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR LINEAR
;  TITLE OF INVENTION:  ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES
;  FILE REFERENCE:  492692000100
;  CURRENT APPLICATION NUMBER: US/09/660,877
;  CURRENT FILING DATE:  2000-09-13
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Synthetic primer
US-09-660-877-17

Alignment Scores:
Pred. No.:              30.8           Length:        351
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            2.45%          Indels:        0
DB:                     4              Gaps:          0

US-09-357-675C-21 (1-327) x US-09-660-877-17 (1-351)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 5
US-09-439-261-12/c
; Sequence 12, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
;  APPLICANT:  Abbott Laboratories
;  APPLICANT:  Mukerji, Pardip
;  APPLICANT:  Leonard, Amanda E.
;  APPLICANT:  Huang, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P2
;  CURRENT APPLICATION NUMBER: US/09/439,261
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  US 08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  PRIOR APPLICATION NUMBER:  PCT/US98/07422
;  PRIOR FILING DATE:  1998-04-10
;  PRIOR APPLICATION NUMBER:  US 09/227,613
;  PRIOR FILING DATE:  1999-01-08
;  NUMBER OF SEQ ID NOS:  60
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-439-261-12

Alignment Scores:
Pred. No.:              71.9           Length:        864
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            2.45%          Indels:        0
DB:                     4              Gaps:          0

US-09-357-675C-21 (1-327) x US-09-439-261-12 (1-864)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    361 AGAGAGGCAGCAAGGCTGGGGGCC 338

RESULT 6
US-09-227-613-13/c
; Sequence 13, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
;  APPLICANT:  MUKERJI, Pradip
;  APPLICANT:  LEONARD, Amanda E.
;  APPLICANT:  HUANG, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P1
;  CURRENT APPLICATION NUMBER: US/09/227,613A
;  CURRENT FILING DATE:  1999-01-08
;  PRIOR APPLICATION NUMBER:  08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 864
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-227-613-13

Alignment Scores:
Pred. No.:              71.9           Length:        864
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            2.45%          Indels:        0
DB:                     4              Gaps:          0

US-09-357-675C-21 (1-327) x US-09-227-613-13 (1-864)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    361 AGAGAGGCAGCAAGGCTGGGGGCC 338

RESULT 7
US-09-439-261-5/c
; Sequence 5, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
;  APPLICANT:  Abbott Laboratories
;  APPLICANT:  Mukerji, Pardip
;  APPLICANT:  Leonard, Amanda E.
;  APPLICANT:  Huang, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P2
;  CURRENT APPLICATION NUMBER: US/09/439,261
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  US 08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  PRIOR APPLICATION NUMBER:  PCT/US98/07422
;  PRIOR FILING DATE:  1998-04-10
;  PRIOR APPLICATION NUMBER:  US 09/227,613
;  PRIOR FILING DATE:  1999-01-08
;  NUMBER OF SEQ ID NOS:  60
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 918
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (755)...(755)
;   OTHER INFORMATION: r = g or a at position 755
US-09-439-261-5

Alignment Scores:
Pred. No.:              76.1          Length:         918
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-439-261-5 (1-918)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    796 AGAGAGGCAGCAAGGCTGGGGGCC 773

RESULT 8
US-09-227-613-5/c
; Sequence 5, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
;  APPLICANT:  MUKERJI, Pradip
;  APPLICANT:  LEONARD, Amanda E.
;  APPLICANT:  HUANG, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P1
;  CURRENT APPLICATION NUMBER: US/09/227,613A
;  CURRENT FILING DATE:  1999-01-08
;  PRIOR APPLICATION NUMBER:  08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 918
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-227-613-5

Alignment Scores:
Pred. No.:              76.1          Length:         918
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-227-613-5 (1-918)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    796 AGAGAGGCAGCAAGGCTGGGGGCC 773

RESULT 9
US-09-439-261-36/c
; Sequence 36, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
;  APPLICANT:  Abbott Laboratories
;  APPLICANT:  Mukerji, Pardip
;  APPLICANT:  Leonard, Amanda E.
;  APPLICANT:  Huang, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P2
;  CURRENT APPLICATION NUMBER: US/09/439,261
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  US 08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  PRIOR APPLICATION NUMBER:  PCT/US98/07422
;  PRIOR FILING DATE:  1998-04-10
;  PRIOR APPLICATION NUMBER:  US 09/227,613
;  PRIOR FILING DATE:  1999-01-08
;  NUMBER OF SEQ ID NOS:  60
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-439-261-36

Alignment Scores:
Pred. No.:              79.4          Length:         960
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-439-261-36 (1-960)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    862 AGAGAGGCAGCAAGGCTGGGGGCC 839

RESULT 10
US-09-227-613-35/c
; Sequence 35, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
;  APPLICANT:  MUKERJI, Pradip
;  APPLICANT:  LEONARD, Amanda E.
;  APPLICANT:  HUANG, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P1
;  CURRENT APPLICATION NUMBER: US/09/227,613A
;  CURRENT FILING DATE:  1999-01-08
;  PRIOR APPLICATION NUMBER:  08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-227-613-35

Alignment Scores:
Pred. No.:              79.4          Length:         960
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-227-613-35 (1-960)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    862 AGAGAGGCAGCAAGGCTGGGGGCC 839

RESULT 11
US-09-439-261-35/c
; Sequence 35, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
;  APPLICANT:  Abbott Laboratories
;  APPLICANT:  Mukerji, Pardip
;  APPLICANT:  Leonard, Amanda E.
;  APPLICANT:  Huang, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P2
;  CURRENT APPLICATION NUMBER: US/09/439,261
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  US 08/833,610
;  PRIOR FILING DATE:  1997-04-11
```

```
;  PRIOR APPLICATION NUMBER:  PCT/US98/07422
;  PRIOR FILING DATE:  1998-04-10
;  PRIOR APPLICATION NUMBER:  US 09/227,613
;  PRIOR FILING DATE:  1999-01-08
;  NUMBER OF SEQ ID NOS:  60
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-439-261-35

Alignment Scores:
Pred. No.:              81.7          Length:         990
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-439-261-35 (1-990)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    832 AGAGAGGCAGCAAGGCTGGGGGCC 809

RESULT 12
US-09-227-613-34/c
; Sequence 34, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
;  APPLICANT:  MUKERJI, Pradip
;  APPLICANT:  LEONARD, Amanda E.
;  APPLICANT:  HUANG, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P1
;  CURRENT APPLICATION NUMBER: US/09/227,613A
;  CURRENT FILING DATE:  1999-01-08
;  PRIOR APPLICATION NUMBER:  08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-227-613-34

Alignment Scores:
Pred. No.:              81.7          Length:         990
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-227-613-34 (1-990)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    832 AGAGAGGCAGCAAGGCTGGGGGCC 809

RESULT 13
US-09-439-261-1/c
; Sequence 1, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
;  APPLICANT:  Abbott Laboratories
;  APPLICANT:  Mukerji, Pardip
;  APPLICANT:  Leonard, Amanda E.
;  APPLICANT:  Huang, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P2
;  CURRENT APPLICATION NUMBER: US/09/439,261
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  US 08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  PRIOR APPLICATION NUMBER:  PCT/US98/07422
;  PRIOR FILING DATE:  1998-04-10
;  PRIOR APPLICATION NUMBER:  US 09/227,613
;  PRIOR FILING DATE:  1999-01-08
;  NUMBER OF SEQ ID NOS:  60
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1335
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-439-261-1

Alignment Scores:
Pred. No.:              108           Length:         1335
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-439-261-1 (1-1335)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    832 AGAGAGGCAGCAAGGCTGGGGGCC 809

RESULT 14
US-09-227-613-1/c
; Sequence 1, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
;  APPLICANT:  MUKERJI, Pradip
;  APPLICANT:  LEONARD, Amanda E.
;  APPLICANT:  HUANG, Yung-Sheng
;  TITLE OF INVENTION:  HUMAN DESATURASE GENE AND USES THEREOF
;  FILE REFERENCE:  6295.US.P1
;  CURRENT APPLICATION NUMBER: US/09/227,613A
;  CURRENT FILING DATE:  1999-01-08
;  PRIOR APPLICATION NUMBER:  08/833,610
;  PRIOR FILING DATE:  1997-04-11
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1335
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-227-613-1

Alignment Scores:
Pred. No.:              108           Length:         1335
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     4             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-227-613-1 (1-1335)

Qy     72 ArgGluAlaAlaArgLeuGlyAla 79
          |||||||||||||||||||||||
Db    832 AGAGAGGCAGCAAGGCTGGGGGCC 809

RESULT 15
US-09-039-046-1/c
; Sequence 1, Application US/09039046
; Patent No. 6331660
; GENERAL INFORMATION:
;  APPLICANT:  CHOMET, PAUL S.
;  APPLICANT:  FREY, MONIKA
```

```
;  APPLICANT:  GIERL, ALFONS
;  TITLE OF INVENTION:  MAIZE DIMBOA BIOSYNTHESIS GENES
;  FILE REFERENCE:  DKEM:131
;  CURRENT APPLICATION NUMBER: US/09/039,046
;  CURRENT FILING DATE:  1998-03-13
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1508
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (49)..(1089)
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Primer
US-09-039-046-1

Alignment Scores:
Pred. No.:              121          Length:          1508
Score:                  8.00         Matches:         8
Percent Similarity:     100.00%      Conservative:    0
Best Local Similarity:  100.00%      Mismatches:      0
Query Match:            2.45%        Indels:          0
DB:                     4            Gaps:            0

US-09-357-675C-21 (1-327) x US-09-039-046-1 (1-1508)

Qy     302 ArgGlnLeuArgArgHisLeuPro 309
           ||||||||||||||||||||||||
Db     585 CGACAGCTCCGGCGTCACCTCCCT 562


Search completed: December  8, 2002, 12:02:50
Job time : 56 secs
```

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2002, 10:39:26 ; Search time 58 Seconds
(without alignments)
2198.414 Million cell updates/sec

Title: US-09-357-675C-21
Perfect score: 327
Sequence: 1 MLGFITRPPHRFLSLLCPGL.....LPVFQHRRPDLYGNLGHPLS 327

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 350425 seqs, 194966369 residues

Word size: 1

Total number of hits satisfying chosen parameters: 699917

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09357675/runat_03122002_142237_20448/app_query.fasta_1.519
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09357675_@CGN_1_1_21_@runat_03122002_142237_20448
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	253	77.4	1382	10	US-09-925-300-692	Sequence 692, App
c	2	63	19.3	377	9	US-09-736-457-1096	Sequence 1096, Ap
c	3	63	19.3	377	9	US-09-902-941-1096	Sequence 1096, Ap
	4	13	4.0	604	10	US-09-770-149-929	Sequence 929, App
	5	8	2.4	56	10	US-09-866-514-5	Sequence 5, Appli
c	6	8	2.4	115	10	US-09-870-433-18	Sequence 18, Appl
c	7	8	2.4	115	10	US-09-893-191-3	Sequence 3, Appli
	8	8	2.4	120	10	US-09-969-373-1197	Sequence 1197, Ap
	9	8	2.4	198	10	US-09-969-373-1198	Sequence 1198, Ap
c	10	8	2.4	250	10	US-09-764-847-1703	Sequence 1703, Ap
c	11	8	2.4	260	10	US-09-878-574-10410	Sequence 10410, A
c	12	8	2.4	334	10	US-09-867-550-1983	Sequence 1983, Ap
c	13	8	2.4	351	10	US-09-870-433-17	Sequence 17, Appl
c	14	8	2.4	351	10	US-09-893-191-4	Sequence 4, Appli
c	15	8	2.4	380	10	US-09-917-800A-1047	Sequence 1047, Ap
	16	8	2.4	762	10	US-09-910-943-678	Sequence 678, App
c	17	8	2.4	1042	10	US-09-844-493-12	Sequence 12, Appl
c	18	8	2.4	1042	10	US-09-844-501-12	Sequence 12, Appl
c	19	8	2.4	1042	10	US-09-844-265-12	Sequence 12, Appl
c	20	8	2.4	1107	9	US-09-764-868-539	Sequence 539, App
	21	8	2.4	1896	10	US-09-822-849A-255	Sequence 255, App
	22	8	2.4	1972	10	US-09-822-849A-485	Sequence 485, App
	23	8	2.4	2210	12	US-10-052-586-287	Sequence 287, App
	24	8	2.4	2212	9	US-09-764-868-540	Sequence 540, App
c	25	8	2.4	2217	9	US-09-764-868-166	Sequence 166, App
c	26	8	2.4	2324	10	US-09-917-800A-1623	Sequence 1623, Ap
c	27	8	2.4	3139	10	US-09-960-428-14	Sequence 14, Appl
c	28	8	2.4	3300	10	US-09-379-931-6	Sequence 6, Appli
	29	8	2.4	3350	10	US-09-880-107-2264	Sequence 2264, Ap
c	30	8	2.4	5011	10	US-09-764-847-1562	Sequence 1562, Ap
	31	8	2.4	6284	10	US-09-764-877-3643	Sequence 3643, Ap
c	32	8	2.4	33239	10	US-09-814-950-3	Sequence 3, Appli
c	33	8	2.4	45839	12	US-10-025-187-3	Sequence 3, Appli
	34	8	2.4	58837	10	US-09-982-091A-5	Sequence 5, Appli
c	35	8	2.4	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c	36	7	2.1	27	10	US-09-374-671-75	Sequence 75, Appl
	37	7	2.1	43	10	US-09-741-849-3	Sequence 3, Appli
	38	7	2.1	48	10	US-09-865-578-9	Sequence 9, Appli
	39	7	2.1	174	10	US-09-783-590-5923	Sequence 5923, Ap
c	40	7	2.1	174	10	US-09-974-300-4734	Sequence 4734, Ap
	41	7	2.1	181	10	US-09-865-578-10	Sequence 10, Appl
c	42	7	2.1	202	10	US-09-867-701-8310	Sequence 8310, Ap
c	43	7	2.1	232	10	US-09-864-761-22886	Sequence 22886, A
c	44	7	2.1	235	10	US-09-923-876-3099	Sequence 3099, Ap
c	45	7	2.1	239	10	US-09-777-564-242	Sequence 242, App

ALIGNMENTS

RESULT 1
US-09-925-300-692
; Sequence 692, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
; LENGTH: 1382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-692

Alignment Scores:

Pred. No.:	1.11e-243	Length:	1382
Score:	253.00	Matches:	326
Percent Similarity:	99.69%	Conservative:	0
Best Local Similarity:	99.69%	Mismatches:	1

```
Query Match:            77.37%      Indels:        1
DB:                     10          Gaps:          0

US-09-357-675C-21 (1-327) x US-09-925-300-692 (1-1382)

Qy       1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeu 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 ATGCTGGGCTTCATCACCAGGCCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTC 136

Qy      21 ArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 CGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCC 196

Qy      41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLys 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 TCTTCCTCCTGCGAACTGCCCCTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAG 256

Qy      61 GlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CAACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGC 316

Qy      81 LeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTGGCTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACAC 376

Qy     101 LeuSerGluProLeuGlyGlyLysLeuGluGluTyrThrGlnLeuAlaArgGluCys 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CTGTCTGAACCACTGGGTGGGAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGT 436

Qy     121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 GGACTCTGGCTGTCCTTGGGTGGTTTCCATGAGCGTGGCCAAGACTGGGAGCAGACTCAG 496

Qy     141 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 AAAATCTACAATTGTCACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCACTTACAGG 556

Qy     161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
           ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db     557 AAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCT-ATGTGTGAAAGCAACTCT 615

Qy     181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     616 ACCATGCCTGGGCCCAGTCTTGAGTCACCTGTCAGCACACCAGCAGGCAAGATTGGTCTA 675

Qy     201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     676 GCTGTCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGGCATTGGCTCAAGCTGGAGCA 735

Qy     221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     736 GAGATACTTACCTATCCTTCAGCTTTTGGATCCATTACAGGCCCAGCCCACTGGGAGGTG 795

Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGly 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     796 TTGCTGCGGGCCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTGTGGA 855

Qy     261 ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     856 CGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGGGAACA 915

Qy     281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     916 GTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCTCTGCCTTGCCCGAATAGACCTCAACTAT 975

Qy     301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     976 CTGCGACAGTTGCGCCGACACCTGCCTGTGTTCCAGCACCGCAGGCCTGACCTCTATGGC 1035

Qy     321 AsnLeuGlyHisProLeuSer 327
           |||||||||||||||||||
Db    1036 AATCTGGGTCACCCACTGTCT 1056
```

```
RESULT 2
US-09-736-457-1096/c
; Sequence 1096, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-736-457-1096

Alignment Scores:
Pred. No.:              2.43e-54      Length:         377
Score:                  63.00         Matches:        124
Percent Similarity:     98.41%        Conservative:   0
Best Local Similarity:  98.41%        Mismatches:     1
Query Match:            19.27%        Indels:         2
DB:                     9             Gaps:           0

US-09-357-675C-21 (1-327) x US-09-736-457-1096 (1-377)

Qy       8 ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 27
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy      28 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuPro 47
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy      48 LeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy      68 AlaGlu-LeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAl 87
           |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGG-GCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGC 139

Qy      87 aPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGl 107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGG 79

Qy     107 yLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGG 19

Qy     127 yGlyPheHisGluArg 132
           ||||||||||||||||
Db      18 TGGTTTCCATGAGCGT 3

RESULT 3
US-09-902-941-1096/c
; Sequence 1096, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
```

```
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1096
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-902-941-1096

Alignment Scores:
Pred. No.:              2.43e-54        Length:         377
Score:                  63.00           Matches:        124
Percent Similarity:     98.41%          Conservative:   0
Best Local Similarity:  98.41%          Mismatches:     1
Query Match:            19.27%          Indels:         2
DB:                     9               Gaps:           0

US-09-357-675C-21 (1-327) x US-09-902-941-1096 (1-377)

Qy       8 ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 27
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 CCTCCTCACAGATTCCTGTCCCTTCTGTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318

Qy      28 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuPro 47
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 CTTTGTGCTCAGCCCAGGCCCAGAGCCATGGCTATCTCCTCTTCCTCCTGCGAACTGCCC 258

Qy      48 LeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 CTGGTGGCTGTGTGCCAGGTAACATCGACGCCAGACAAGCAACAGAACTTTAAAACATGT 198

Qy      68 AlaGlu-LeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAl 87
           |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 GCTGG-GCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTCCTGCCTGAGGC 139

Qy      87 aPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGl 107
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ATTTGACTTCATTGCACGGGACCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGG 79

Qy     107 yLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GAAACTTTTGGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTGTCCTTGGG 19

Qy     127 yGlyPheHisGluArg 132
           |||||||||||||||
Db      18 TGGTTTCCATGAGCGT 3

RESULT 4
US-09-770-149-929
; Sequence 929, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
;  APPLICANT:  Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Rameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION:  Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE:  2024 (PARA-013PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,149
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER:  60/178,506
;  PRIOR FILING DATE:  2000-01-27
;  NUMBER OF SEQ ID NOS:  999
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 929
;   LENGTH: 604
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(604)
;   OTHER INFORMATION: n = A,T,C or G
US-09-770-149-929

Alignment Scores:
Pred. No.:              0.000345        Length:         604
Score:                  13.00           Matches:        13
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            3.98%           Indels:         0
DB:                     10              Gaps:           0

US-09-357-675C-21 (1-327) x US-09-770-149-929 (1-604)

Qy     241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 253
           |||||||||||||||||||||||||||||||||||||
Db     198 CTTCTTCGAGCCCGAGCAATTGAAACTCAATGTTATGTC 236

RESULT 5
US-09-866-514-5
; Sequence 5, Application US/09866514
; Patent No. US20020137050A1
; GENERAL INFORMATION:
;  APPLICANT: Elsbach and Weiss
;  TITLE OF INVENTION: THERAPEUTIC USES OF BACTERICIDAL/PERMEABILITY-INCREASING
;  TITLE OF INVENTION:  PROTEIN FRAGMENTS
;  FILE REFERENCE: 28297/32248B
;  CURRENT APPLICATION NUMBER: US/09/866,514
;  CURRENT FILING DATE:  2001-05-25
;  PRIOR APPLICATION NUMBER: 09/309,217
;  PRIOR FILING DATE: 1999-05-10
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 56
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Primer
US-09-866-514-5

Alignment Scores:
Pred. No.:              3.82            Length:         56
Score:                  8.00            Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            2.45%           Indels:         0
```

```
DB:                    10            Gaps:          0

US-09-357-675C-21 (1-327) x US-09-866-514-5 (1-56)

Qy    286 SerGluGlyProGlyLeuCysLeu 293
          |||||||||||||||||||||||
Db     30 TCAGAAGGGCCTGGATTATGCCTC 53

RESULT 6
US-09-870-433-18/c
; Sequence 18, Application US/09870433
; Patent No. US20010034048A1
; GENERAL INFORMATION:
;  APPLICANT: Nurith KURN
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR
;  TITLE OF INVENTION:  ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES
;  FILE REFERENCE: 492692000101
;  CURRENT APPLICATION NUMBER: US/09/870,433
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: 09/660,877
;  PRIOR FILING DATE: 2000-09-13
;  PRIOR APPLICATION NUMBER: 60/175,780
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/153,604
;  PRIOR FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 115
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: IA013
US-09-870-433-18

Alignment Scores:
Pred. No.:             7.41          Length:        115
Score:                 8.00          Matches:       8
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           2.45%         Indels:        0
DB:                    10            Gaps:          0

US-09-357-675C-21 (1-327) x US-09-870-433-18 (1-115)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 7
US-09-893-191-3/c
; Sequence 3, Application US/09893191
; Patent No. US20020058270A1
; GENERAL INFORMATION:
;  APPLICANT: Kurn, Nurith
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;  TITLE OF INVENTION:  TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION
;  FILE REFERENCE: 49269-20006.00
;  CURRENT APPLICATION NUMBER: US/09/893,191
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/213,908
;  PRIOR FILING DATE: 2000-06-26
;  PRIOR APPLICATION NUMBER: 60/277,748
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 115
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Primer: IA013
US-09-893-191-3

Alignment Scores:
Pred. No.:             7.41          Length:        115
Score:                 8.00          Matches:       8
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           2.45%         Indels:        0
DB:                    10            Gaps:          0

US-09-357-675C-21 (1-327) x US-09-893-191-3 (1-115)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 8
US-09-969-373-1197
; Sequence 1197, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1197
;   LENGTH: 120
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-1197

Alignment Scores:
Pred. No.:             7.71          Length:        120
Score:                 8.00          Matches:       8
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           2.45%         Indels:        0
DB:                    10            Gaps:          0

US-09-357-675C-21 (1-327) x US-09-969-373-1197 (1-120)

Qy      9 ProHisArgPheLeuSerLeuLeu 16
          |||||||||||||||||||||||
Db     36 CCTCATAGATTCCTATCCCTCCTT 59

RESULT 9
US-09-969-373-1198
; Sequence 1198, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1198
;   LENGTH: 198
```

```
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-1198

Alignment Scores:
Pred. No.:              12.2          Length:         198
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-969-373-1198 (1-198)

Qy      9 ProHisArgPheLeuSerLeuLeu 16
          ||||||||||||||||||||||||
Db     14 CCTCATAGATTCCTATCCCTCCTT 37

RESULT 10
US-09-764-847-1703/c
; Sequence 1703, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1703
;   LENGTH: 250
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1703

Alignment Scores:
Pred. No.:              15.2          Length:         250
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-764-847-1703 (1-250)

Qy    212 SerLeuAlaLeuAlaGlnAlaGly 219
          |||||||||||||||||||||||
Db    225 AGTCTCGCTCTTGCCCAGGCTGGA 202

RESULT 11
US-09-878-574-10410/c
; Sequence 10410, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10410
;   LENGTH: 260
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701103644H1
US-09-878-574-10410

Alignment Scores:
Pred. No.:              15.7          Length:         260
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-878-574-10410 (1-260)

Qy     39 IleSerSerSerSerCysGluLeu 46
          |||||||||||||||||||||||||
Db     66 ATCTCTTCCTCTTCCTGTGAACTC 43

RESULT 12
US-09-867-550-1983/c
; Sequence 1983, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells a
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1983
;   LENGTH: 334
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (334)
;   OTHER INFORMATION: Wherein n is one of a or t or c or g
US-09-867-550-1983

Alignment Scores:
Pred. No.:              19.8          Length:         334
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-867-550-1983 (1-334)

Qy      9 ProHisArgPheLeuSerLeuLeu 16
          ||||||||||||||||||||||||
Db     27 CCACACCGATTTCTTTCTTTGCTC 4

RESULT 13
US-09-870-433-17/c
; Sequence 17, Application US/09870433
; Patent No. US20010034048A1
; GENERAL INFORMATION:
;  APPLICANT: Nurith KURN
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR
;  TITLE OF INVENTION:  ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES
;  FILE REFERENCE: 492692000101
;  CURRENT APPLICATION NUMBER: US/09/870,433
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: 09/660,877
;  PRIOR FILING DATE: 2000-09-13
;  PRIOR APPLICATION NUMBER: 60/175,780
;  PRIOR FILING DATE: 2000-01-12
```

```
;  PRIOR APPLICATION NUMBER: 60/153,604
;  PRIOR FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR Primer
US-09-870-433-17

Alignment Scores:
Pred. No.:              20.7          Length:         351
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-870-433-17 (1-351)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 14
US-09-893-191-4/c
; Sequence 4, Application US/09893191
; Patent No. US20020058270A1
; GENERAL INFORMATION:
;  APPLICANT: Kurn, Nurith
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;  TITLE OF INVENTION:  TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION
;  FILE REFERENCE: 49269-20006.00
;  CURRENT APPLICATION NUMBER: US/09/893,191
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/213,908
;  PRIOR FILING DATE: 2000-06-26
;  PRIOR APPLICATION NUMBER: 60/277,748
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Primer
US-09-893-191-4

Alignment Scores:
Pred. No.:              20.7          Length:         351
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-893-191-4 (1-351)

Qy    189 SerProValSerThrProAlaGly 196
          |||||||||||||||||||||||||
Db    105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 15
US-09-917-800A-1047/c
; Sequence 1047, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1047
;   LENGTH: 380
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI178850
US-09-917-800A-1047

Alignment Scores:
Pred. No.:              22.3          Length:         380
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            2.45%         Indels:         0
DB:                     10            Gaps:           0

US-09-357-675C-21 (1-327) x US-09-917-800A-1047 (1-380)

Qy    282 ValAlaArgCysSerGluGlyPro 289
          |||||||||||||||||||||||
Db    194 GTGGCAAGGTGTAGTGAGGGACCT 171


Search completed: December  8, 2002, 12:04:09
Job time : 61 secs
```